

L Number	Hits	Search Text	DB	Time stamp
7	1	wo NEAR "9605306"	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/27 16:52
-	2	("5654155").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/23 13:00
-	1	("20020183268").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:03
-	9	(Murphy NEAR Patricia) and BRCA1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:07
-	170	BRCA1 WITH sequence	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:09
-	164	(BRCA1 WITH sequence) and gene	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:08
-	12	((BRCA1 WITH sequence) and gene) and omi\$1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:08
-	8	BRCA1 SAME sequence SAME gene SAME omi\$1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/01/14 17:09
-	116	Stommel	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/02 14:14
-	1	Stommel and brca1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/02 14:15
-	700	Durocher	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/02 14:15
-	10	Durocher and brca1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/02 14:15
-	20	holt NEAR jeffrey	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/18 12:41
-	11	(holt NEAR jeffrey) and brca1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/18 12:41
-	962	BRCA1	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/18 12:51
-	13	BRCA1 and (thymidine WITH cytidine)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/18 12:52
-	28	Murphy NEAR Patricia	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/19 10:46





-	6	(Murphy NEAR Patricia) and SEQ.clm.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/19 10:46
-	49	((("5547839") or ("5510270") or ("6045997") or ("5561058") or ("5455934") or ("5891857") or ("6051379") or ("5858669") or ("4683202") or ("6130322") or ("5750400") or ("5911227") or ("5624803") or ("6083698") or ("5545531") or ("5948643") or ("5693473") or ("5589330") or ("5633134") or ("5726019") or ("5710001") or ("5753441") or ("5747282") or ("6033857") or ("6124104") or ("5756294")).PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/23 13:15
-	2	("5912127").PN.	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/06/23 13:16



```
Db 301 GTCCCTTATGTAAGATGATTAACCAAAAAGAGCCTACAAGAAATGACAGATTTAGTC 360
QY 361 AACTGTGTAAGACCTAATGAAATCATTTTGTGTTTTCAGCTTGAACAGGTTTGGAGT 420
Db 361 AACTGTGTAAGACCTAATGAAATCATTTTGTGTTTTCAGCTTGAACAGGTTTGGAGT 420
QY 421 ATGCAACAGCTAATATTTTGCAAAAGAAAATACTCTCCGGAACATCTAAAGATG 480
Db 421 ATGCAACAGCTAATATTTTGCAAAAGAAAATACTCTCCGGAACATCTAAAGATG 480
QY 481 AAGTTTCTATCACCAGATGATGGGCTACAGAAAACGTGCCAAAAGACTTCTACAGAGT 540
Db 481 AAGTTTCTATCACCAGATGATGGGCTACAGAAAACGTGCCAAAAGACTTCTACAGAGT 540
QY 541 AACCAGAAATCCTCTCTGAGGAAACGATCCTCACTGTCACACTCTCACTCACTTGGAA 600
Db 541 AACCAGAAATCCTCTCTGAGGAAACGATCCTCACTGTCACACTCTCACTCACTTGGAA 600
QY 601 CTGTGAGACCTCTGAGGACAAAGCAGCGGATACAACTCAAAAGACGTCGTCTACATG 660
Db 601 CTGTGAGACCTCTGAGGACAAAGCAGCGGATACAACTCAAAAGACGTCGTCTACATG 660
QY 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGCACTTATTCAGTGTGGAG 720
Db 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGCACTTATTCAGTGTGGAG 720
QY 721 ATCAAGATTTGTTACAAANTACCCCTCAGAGAACCGAGATGAATAGTTGGATTCTG 780
Db 721 ATCAAGATTTGTTACAAANTACCCCTCAGAGAACCGAGATGAATAGTTGGATTCTG 780
QY 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGCGATGTACAAATACTGAACATCATCAAC 840
Db 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGCGATGTACAAATACTGAACATCATCAAC 840
QY 841 CCACTAATTAATGATTTTAACACACACAGAGGCTGAGAGGCTCCAGAAAAGT 900
Db 841 CCACTAATTAATGATTTTAACACACACAGAGGCTGAGAGGCTCCAGAAAAGT 900
QY 901 ATCAGGATGATCTGTTTCAAACTGATGATGAGGACATGAGCAAAATACATCAATGCCA 960
Db 901 ATCAGGATGATCTGTTTCAAACTGATGATGAGGACATGAGCAAAATACATCAATGCCA 960
QY 961 GCTCATTAACAGCATGAGAACAGCAGTTTATTACTCACTAAGACAGAAATGATGAGAA 1020
Db 961 GCTCATTAACAGCATGAGAACAGCAGTTTATTACTCACTAAGACAGAAATGATGAGAA 1020
QY 1021 AGGCTGAATCTGTAATAAAGCAAAACAGCTGCTTACGAGAGGCAACATTAACAGAT 1080
Db 1021 AGGCTGAATCTGTAATAAAGCAAAACAGCTGCTTACGAGAGGCAACATTAACAGAT 1080
QY 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGACCTCCACACAGAAAAGGTTAG 1140
Db 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGACCTCCACACAGAAAAGGTTAG 1140
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGGAATAGCAGAAATCTCCATGCT 1200
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGGAATAGCAGAAATCTCCATGCT 1200
QY 1201 CAGAGAAATCCTAGAGATGATGAGATGTTCTCTGATACACATAAATAGCAGATTTCAGA 1260
Db 1201 CAGAGAAATCCTAGAGATGATGAGATGTTCTCTGATACACATAAATAGCAGATTTCAGA 1260
QY 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTATGATGATGTTAGTTCTGATGATCACTAGATG 1320
Db 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTATGATGATGTTAGTTCTGATGATCACTAGATG 1320
QY 1321 GGGAGTCTGAATCAAAATGCAAAAGTACGTATGATGACGTTCTAATAGAGTATGATG 1380
Db 1321 GGGAGTCTGAATCAAAATGCAAAAGTACGTATGATGACGTTCTAATAGAGTATGATG 1380
QY 1381 AATATTTCTGTTCTTCAAGAAATAGACTTACTGGCCAGTGTCTCTCATGAGAGCTTTAA 1440
Db 1381 AATATTTCTGTTCTTCAAGAAATAGACTTACTGGCCAGTGTCTCTCATGAGAGCTTTAA 1440
QY 1441 TATGTAAAGTGAAGAGTTCACCTCAATCAGTAGAGATTAATTTGAAGACAAATAT 1500
Db 1441 TATGTAAAGTGAAGAGTTCACCTCAATCAGTAGAGATTAATTTGAAGACAAATAT 1500
QY 1501 TTGGGAAACCTATGCGAAGAGGCAAGCCTCCCAACTTAAGCAGTATTAAGTGAATTC 1560
Db 1501 TTGGGAAACCTATGCGAAGAGGCAAGCCTCCCAACTTAAGCAGTATTAAGTGAATTC 1560
QY 1561 TAAATATGAGAGATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAATTA 1620
Db 1561 TAAATATGAGAGATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAATTA 1620
QY 1621 AATTAAGGCTAAAGAGAGCTACATCAGAGCCTTCATCCTGAGATTTTATTAAGAAAG 1680
Db 1621 AATTAAGGCTAAAGAGAGCTACATCAGAGCCTTCATCCTGAGATTTTATTAAGAAAG 1680
QY 1681 CAGATTTGGCAGTTCAAAAACATCTGCAATGATTAATACAGGAGACTAACCAACGAGC 1740
Db 1681 CAGATTTGGCAGTTCAAAAACATCTGCAATGATTAATACAGGAGACTAACCAACGAGC 1740
QY 1741 AGAATGCTCAAGTATGATTAATTAATTAATAGTGTCTACAGATTAATAAAGTGTATT 1800
Db 1741 AGAATGCTCAAGTATGATTAATTAATTAATAGTGTCTACAGATTAATAAAGTGTATT 1800
QY 1801 CFATTCAGATGAGAAAATCCTAACCCATAGATACATCGAAAAGAAATCTGCTTTCA 1860
Db 1801 CFATTCAGATGAGAAAATCCTAACCCATAGATACATCGAAAAGAAATCTGCTTTCA 1860
QY 1861 AAGCAAGAGCTGAACCTATTAAGCAGAGATTAAGCAATATGAGAACTGAATTAATATCC 1920
Db 1861 AAGCAAGAGCTGAACCTATTAAGCAGAGATTAAGCAATATGAGAACTGAATTAATATCC 1920
QY 1921 ACAATTTCAAAAGCACCCTAAAGAAATAGGCTGAGAGGAAAGTCTTACAGAGCATATTC 1980
Db 1921 ACAATTTCAAAAGCACCCTAAAGAAATAGGCTGAGAGGAAAGTCTTACAGAGCATATTC 1980
QY 1981 ATGCGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
Db 1981 ATGCGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
QY 2041 TTGATAGTGTGTTTACAGATGAGAGATTAAGAAAAAGTATCAACCAATCCAGTCA 2100
Db 2041 TTGATAGTGTGTTTACAGATGAGAGATTAAGAAAAAGTATCAACCAATCCAGTCA 2100
QY 2101 GGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGAGGCCAAGAGA 2160
Db 2101 GGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGAGGCCAAGAGA 2160
QY 2161 GTAACCAAGCCAAATGAAAGACAGAAATGAAAGACATGACAGTATCTTCCAGAGCTGA 2220
Db 2161 GTAACCAAGCCAAATGAAAGACAGAAATGAAAGACATGACAGTATCTTCCAGAGCTGA 2220
QY 2221 AGTTAACAAATGACACCTGTTCTTTTACTAAGTGTCAATATACAGTGAATTAAGAAAT 2280
Db 2221 AGTTAACAAATGACACCTGTTCTTTTACTAAGTGTCAATATACAGTGAATTAAGAAAT 2280
QY 2281 TTGTCAATCTTACCTTCCAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTGT 2340
Db 2281 TTGTCAATCTTACCTTCCAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTGT 2340
QY 2341 CTATATATGCTGAAGACCCCAAGATCTATGTTAGTGGAGAAAGGTTTGGCAACAG 2400
Db 2341 CTATATATGCTGAAGACCCCAAGATCTATGTTAGTGGAGAAAGGTTTGGCAACAG 2400
QY 2401 AAGATCTGTAGAGATGACATATTTCAATGATGATGATGATGATGATGATGATGATGATG 2460
Db 2401 AAGATCTGTAGAGATGACATATTTCAATGATGATGATGATGATGATGATGATGATGATG 2460
QY 2461 AAGATCTGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520
Db 2461 AAGATCTGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520
```

Chr	Position	Strand	Sequence	Length
Db	4921	+	AAGTCCCAATGGAAGTTGAGAAATCTGGCCAGAGTCCAGCTCTCATCTACTG	4980
Qy	4981	+	ATACTGCTGGGATATATGCAATGCAAGAAAGTGTGACGAGAGAGACCAATTTGACG	5040
Db	4981	+	ATACTGCTGGGATATATGCAATGCAAGAAAGTGTGACGAGAGAGACCAATTTGACG	5040
Qy	5041	+	CTTCAACGAAAAGGCTCAACAAAACAAATGTCATGCTGTGTCTGTGACCCAGAG	5100
Db	5041	+	CTTCAACGAAAAGGCTCAACAAAACAAATGTCATGCTGTGTCTGTGACCCAGAG	5100
Qy	5101	+	AATTTATCTGCTGTACAAAGTTTGGCCAGAAAACACACATCACTTAATCTAATTA	5160
Db	5101	+	AATTTATCTGCTGTACAAAGTTTGGCCAGAAAACACACATCACTTAATCTAATTA	5160
Qy	5161	+	CTGAAGACACTACATCTGTTGTTATGAAGAAACAGATGCTGAGTTGTGTGAACGACAC	5220
Db	5161	+	CTGAAGACACTACATCTGTTGTTATGAAGAAACAGATGCTGAGTTGTGTGAACGACAC	5220
Qy	5221	+	TGAAATATTTCTGGAATTCGGGAGGAAAATGGGTGTTAGCTATTTCTGGGTGACCC	5280
Db	5221	+	TGAAATATTTCTGGAATTCGGGAGGAAAATGGGTGTTAGCTATTTCTGGGTGACCC	5280
Qy	5281	+	AGCTATTAAAGAAAGAAAATGCGATGAGCATGATTTTGAAGTCAGAGAGATGTG	5340
Db	5281	+	AGCTATTAAAGAAAGAAAATGCGATGAGCATGATTTTGAAGTCAGAGAGATGTG	5340
Qy	5341	+	TCATGGAAGAAACCCACCAGGTCCTCAAGGACAGAGAAATCCAGGACGAAAGATCT	5400
Db	5341	+	TCATGGAAGAAACCCACCAGGTCCTCAAGGACAGAGAAATCCAGGACGAAAGATCT	5400
Qy	5401	+	TCAGGGGCTTGAAGATCTGTTGCTATGGGCTTCCACCAATGCTGATCACTTCA	5460
Db	5401	+	TCAGGGGCTTGAAGATCTGTTGCTATGGGCTTCCACCAATGCTGATCACTTCA	5460
Qy	5461	+	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5520
Db	5461	+	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5520
Qy	5521	+	GCACAGGCTGTCACCAATTTGTTGTTGTCAGCAGATGCTTGACAGAGACAATGGCT	5580
Db	5521	+	GCACAGGCTGTCACCAATTTGTTGTTGTCAGCAGATGCTTGACAGAGACAATGGCT	5580
Qy	5581	+	TCCATGCAATTTGGGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	5640
Db	5581	+	TCCATGCAATTTGGGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	5640
Qy	5641	+	GTGTAGCACTATACAGATGTCAGAGAGTGTGACACCTACTATACCCAGATGCCCA	5700
Db	5641	+	GTGTAGCACTATACAGATGTCAGAGAGTGTGACACCTACTATACCCAGATGCCCA	5700
Qy	5701	+	GCCACTACTGA 5711	5700
Db	5701	+	GCCACTACTGA 5711	5700
RESULT 6				
AAV46464				
ID	AAV46464	standard; cDNA; 5711 BP.		
XX	AAV46464;			
XX	18-NOV-1998	(first entry)		
XX	Human BRCA1 omi2 polymorphism #7 cDNA.			
XX	BRCA1; omi1; human; breast and ovarian cancer predisposing gene;			
KW	polymorphism; susceptibility; anti-oncogene; tumor suppressor;			
KW	chromosome 17q; ss.			
XX	Homo sapiens.			
OS				
XX				
RH	key	Location/Qualifiers		

FT	CDS		120..5711	/tag= a	
FT				/product= "BRCA1 omi2 protein"	
FT				4956	
FT	variation			/tag= b	
FT				/note= "This polymorphic variation can be an A or G nucleotide"	
FN			US5750400-A.		
PD			12-MAY-1998.		
PF			12-FEB-1997;	97US-0798691.	
PR			12-FEB-1996;	96US-0598591.	
PR			12-FEB-1997;	97US-0798691.	
PA	(ONCO- ) ONCORMED INC.				
PI	Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;				
PI	Scheller DB, Zeng B;				
DR	WPI; 1998-296774/26.				
XX					
PT	BRCA1 omi gene coding sequences - useful for distinguishing between				
PT	polymorphisms and mutation(s) in the screening for disposition to				
PT	breast or ovarian cancer				
XX					
PS	Claim 2e; Page -: 54pp; English.				
XX					
CC	This sequence encodes a human BRCA1 (breast and ovarian cancer				
CC	predisposing gene) omi2 gene in which a polymorphic variation occurs at				
CC	nucleotide 4956. This sequence and other polymorphic variations of this				
CC	sequence are useful for the identification of an individual who may or				
CC	may not have an increased susceptibility to breast or ovarian cancer.				
CC	The sequences used identify gene changes which are due to polymorphisms,				
CC	rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour				
CC	suppressor) which is involved in genetic inheritance of cancers,				
CC	especially breast and ovarian cancer. It is found at human chromosome				
CC	17q which is known to be linked to cancer susceptibility, especially				
CC	breast cancer. Cells containing a mutation in this gene lose the				
CC	wild-type function of BRCA1 and are more susceptible to cancers.				
CC	NOTE: This sequence does not appear in the specification but has been				
CC	created from the wild type BRCA1 omi2 gene represented in AAIV4649.				
CS					
XQ	Sequence 5711 BP; 1955 A; 1098 C; 1274 G; 1383 T; 1 other;				
	Query Match	100.0%; Score 5709; DB 19; Length 5711;			
	Best Local Similarity 100.0%; Pred No. 0;				
	Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0				
QY	1 AGCTCGCTGAGACTTCCTGTGCACCCGCCACAGAGCTGTGGGGTTTCTCAGATACTGGGCC	60			
Pb	1 AGCTCGCTGAGACTTCCTGTGCACCCGCCACAGAGCTGTGGGGTTTCTCAGATACTGGGCC	60			
QY	61 CTGCGCTCAGAGAGGCGCTTACCCTTGCTGTGGTAAGTTCAATGGAACAAGAAGAA	120			
Pb	61 CTGCGCTCAGAGAGGCGCTTACCCTTGCTGTGGTAAGTTCAATGGAACAAGAAGAA	120			
QY	121 TGGATTATTCGTCTTGCGGTTGAAGAGTACAAAATGTCATTATGCTTGTGACAAAA	180			
Pb	121 TGGATTATTCGTCTTGCGGTTGAAGAGTACAAAATGTCATTATGCTTGTGACAAAA	180			
QY	181 TCCTTAGAGTGTCCCATCTGCTGTGGAATGATCAAGAACCTGTCCCAAGAGTGTAC	240			
Pb	181 TCCTTAGAGTGTCCCATCTGCTGTGGAATGATCAAGAACCTGTCCCAAGAGTGTAC	240			
QY	241 ACAATATTTTGCAAAATTTGCATGCTGAACCTTCTCACAGAGAAAGGCGCTTCACAGT	300			
Pb	241 ACAATATTTTGCAAAATTTGCATGCTGAACCTTCTCACAGAGAAAGGCGCTTCACAGT	300			
OY	301 GTCTTTATGTAGTAATGATATACCAAAGAGCGCTACAAAGAAAGTACGAGATTATGTC	360			

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 12:34:02 ; Search time 14215.3 Seconds

(without alignments)  
11692.029 Million cell updates/sec

Title: US-09-734-672-3

Perfect score: 5711

Sequence: 1 AGCTGCTGAGACTTCTG.....TCCCCACAGCACTACTGA 5711

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: GenEmbl.\*
- 2: gb\_ba.\*
- 3: gb\_hlg.\*
- 4: gb\_in.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_hlg\_hum.\*
- 31: em\_hlg\_inv.\*
- 32: em\_hlg\_other.\*
- 33: em\_hlg\_mus.\*
- 34: em\_hlg\_pin.\*
- 35: em\_hlg\_rod.\*
- 36: em\_hlg\_mam.\*
- 37: em\_hlg\_vtl.\*
- 38: em\_sy.\*
- 39: em\_hlgo\_hum.\*
- 40: em\_hlgo\_mus.\*
- 41: em\_hlgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5709.4	100.0	5711	6 AR007334	AR007334 Sequence
2	5709.4	100.0	5711	6 AR112810	AR112810 Sequence
3	5707.8	99.9	5711	6 AR033056	AR033056 Sequence
4	5707.8	99.9	5711	6 HSU14680	U14680 Homo sapien
5	5707.8	99.9	5712	6 AR070223	AR070223 Sequence
6	5707.8	99.9	5712	6 AR118507	AR118507 Sequence
7	5707.8	99.9	5712	6 AR125601	AR125601 Sequence
8	5707.8	99.9	5712	6 AR184044	AR184044 Sequence
9	5707.8	99.9	5914	6 AR004673	AR004673 Sequence
10	5707.8	99.9	5914	6 AR008159	AR008159 Sequence
11	5707.8	99.9	5914	6 AR136942	AR136942 Sequence
12	5707.8	99.9	5914	6 I76943	I76943 Sequence 1
13	5707.8	99.9	5914	6 I80938	I80938 Sequence 1
14	5707.8	99.9	5914	6 I81034	I81034 Sequence 1
15	5706.2	99.9	5711	6 AR048660	AR048660 Sequence
16	5706.2	99.9	5711	6 AR048666	AR048666 Sequence
17	5706.2	99.9	5711	6 I40795	I40795 Sequence 4
18	5706.2	99.9	5711	6 I40801	I40801 Sequence 4
19	5704.6	99.9	5711	6 AR007335	AR007335 Sequence
20	5704.6	99.9	5711	6 AR112809	AR112809 Sequence
21	5703	99.9	5711	6 AR007333	AR007333 Sequence
22	5703	99.9	5711	6 AR112808	AR112808 Sequence
23	5703	99.9	5711	6 I59546	I59546 Sequence 1
24	5696.8	99.8	5712	6 AR048668	AR048668 Sequence 12
25	5696.8	99.8	5712	6 I40803	I40803 Sequence 12
26	5695.8	99.7	5710	6 AR048662	AR048662 Sequence 6
27	5695.8	99.7	5710	6 I40797	I40797 Sequence 6
28	5693.8	99.7	5709	6 AR048658	AR048658 Sequence
29	5693.8	99.7	5709	6 AR048663	AR048663 Sequence
30	5693.8	99.7	5709	6 AR048664	AR048664 Sequence
31	5693.8	99.7	5709	6 AR048665	AR048665 Sequence
32	5693.8	99.7	5709	6 I40793	I40793 Sequence 2
33	5693.8	99.7	5709	6 I40798	I40798 Sequence 7
34	5693.8	99.7	5709	6 I40799	I40799 Sequence 8
35	5693.8	99.7	5709	6 I40800	I40800 Sequence 9
36	5689.8	99.6	5707	6 AR048667	AR048667 Sequence
37	5689.8	99.6	5707	6 I40802	I40802 Sequence 11
38	5657.8	99.1	5709	6 AX281859	AX281859 Sequence
39	5653.8	99.0	5689	6 AR048659	AR048659 Sequence
40	5653.8	99.0	5689	6 I40794	I40794 Sequence 3
41	5627.6	98.5	5770	6 AR048661	AR048661 Sequence
42	5627.6	98.5	5770	6 I40796	I40796 Sequence 5
43	5588.8	97.9	5656	6 AR048657	AR048657 Sequence
44	5588.8	97.9	5656	6 I40792	I40792 Sequence 1
45	5530.8	96.8	5693	9 AF005068	AF005068 Homo sapi

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
AR007334	AR007334	Sequence 3 from patent US 5750400.	AR007334	GI:3966818	Unknown.	Unknown.	Unclassified.	1 (bases 1 to 5711)	Murphy,P.D., Allen,A.C.,	Coding sequences of the human BRCA1 gene	Patent: US 5750400-A 3 12-MAY-1998;
5711 bp	DNA	linear	PAT 04-DEC-1998						Alvares,C.P., Critz,B.S., Olson,S.J.,		
									Scheitler,D.B. and Zeng,B.		

FEATURES Location/Qualifiers  
source 1. 5711 /organism="unknown"  
BASE COUNT 1956 a 1098 c 1274 g 1383 t  
ORIGIN

Query Match 100.0%: Score 5709.4: DB 6: Length 5711:  
Best Local Similarity 100.0%: Pred. No. 0:  
Matches 5710: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 AGCTCGTGAGACTTCTCTGAGACCCGACCCAGGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
DB 1 AGCTCGTGAGACTTCTCTGAGACCCGACCCAGGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
QY 61 CTTGGCTCAGAGAGGCTTACCCCTTCTGCTGGGTAAGTTCAATTGGAACAGAAAGAA 120  
DB 61 CTTGGCTCAGAGAGGCTTACCCCTTCTGCTGGGTAAGTTCAATTGGAACAGAAAGAA 120  
QY 121 TGGATTATCTGCTCTCGCGTTGAAGAGTAAATGTCTTAATGCTATGAGAGAAA 180  
DB 121 TGGATTATCTGCTCTCGCGTTGAAGAGTAAATGTCTTAATGCTATGAGAGAAA 180  
QY 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAAACCTGTCTCCACAAAGTGTACC 240  
DB 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAAACCTGTCTCCACAAAGTGTACC 240  
QY 241 ACATATTTTGCMAATTTTGCATGTGAAACTTCTCAACAGAGAAAGGCTTACACAGT 300  
DB 241 ACATATTTTGCMAATTTTGCATGTGAAACTTCTCAACAGAGAAAGGCTTACACAGT 300  
QY 301 GTCTTTATGTAGAATGATATTAACAAAAGAGCTTACAAAGAAAGTACGAGATTTAGTC 360  
DB 301 GTCTTTATGTAGAATGATATTAACAAAAGAGCTTACAAAGAAAGTACGAGATTTAGTC 360  
QY 361 AACTTGTGAAGAGCTATTAATAATCAATTTGTGCTTTTCAAGCTTACAGACAGGTTTGAAGT 420  
DB 361 AACTTGTGAAGAGCTATTAATAATCAATTTGTGCTTTTCAAGCTTACAGACAGGTTTGAAGT 420  
QY 421 ATGCAAAAGCTTAATTTTGGCAAAAAGAAATTAATCTCTCCGAAACATCTAAAGATG 480  
DB 421 ATGCAAAAGCTTAATTTTGGCAAAAAGAAATTAATCTCTCCGAAACATCTAAAGATG 480  
QY 481 AAGTTTCTATCATCAAAAGATGGGTACAGAAAACCTGCAAAAGACTTCTACAGAGTG 540  
DB 481 AAGTTTCTATCATCAAAAGATGGGTACAGAAAACCTGCAAAAGACTTCTACAGAGTG 540  
QY 541 AACCCGAAAAATCTCTCTGAGAGAAACAGTCTCAAGTGTCAAACTCTTAACCTTGCAA 600  
DB 541 AACCCGAAAAATCTCTCTGAGAGAAACAGTCTCAAGTGTCAAACTCTTAACCTTGCAA 600  
QY 601 CTGTGAGACTGTGAGAGCAAAAGCAGGAGTACAACTCAAAAGAGCTGTCTACATTG 660  
DB 601 CTGTGAGACTGTGAGAGCAAAAGCAGGAGTACAACTCAAAAGAGCTGTCTACATTG 660  
QY 661 AATTTGGATCTGATTTCTTCTGAGAGATACGTTAATAAGCAACTTATTTGAGTGGAG 720  
DB 661 AATTTGGATCTGATTTCTTCTGAGAGATACGTTAATAAGCAACTTATTTGAGTGGAG 720  
QY 721 ATCAAGAAATTTGTACAATACACCCTCAAGAAACAGAGATGAATCAGTTTGGATTCTG 780  
DB 721 ATCAAGAAATTTGTACAATACACCCTCAAGAAACAGAGATGAATCAGTTTGGATTCTG 780  
QY 781 CAAAAAAGGCTGCTTGAATTTTCTGAGACGAGATGTAACTACTGAACATCATCAAC 840  
DB 781 CAAAAAAGGCTGCTTGAATTTTCTGAGACGAGATGTAACTACTGAACATCATCAAC 840  
QY 841 CGAGTAATATGATTTGAACACACCTGAGAGAGCTGAGAGGCAATCCAGAAAGT 900  
DB 841 CGAGTAATATGATTTGAACACACCTGAGAGAGCTGAGAGGCAATCCAGAAAGT 900  
QY 901 ATCAGAGGTAGTTCTGTTTCAAACTTGATGTGAGGCAATGTGACACAAATACTCATGCCA 960  
DB 901 ATCAGAGGTAGTTCTGTTTCAAACTTGATGTGAGGCAATGTGACACAAATACTCATGCCA 960

QY 961 GCTATTTACAGATGAGACACGACTTTATTTACTCACTAAAGACAGATGATGAGAA 1020  
DB 961 GCTATTTACAGATGAGACACGACTTTATTTACTCACTAAAGACAGATGATGAGAA 1020  
QY 1021 AGGCTGAATCTGTATATAAAGCAACAGGCTGGTTTACGAGAGGCCCAATACAGAT 1080  
DB 1021 AGGCTGAATCTGTATATAAAGCAACAGGCTGGTTTACGAGAGGCCCAATACAGAT 1080  
QY 1081 GGGCTGGAAGTAAAGAAACATGTATATGATGGGGAGCTCCACACAGAAAAAAGGTAG 1140  
DB 1081 GGGCTGGAAGTAAAGAAACATGTATATGATGGGGAGCTCCACACAGAAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATAGCAGAAACCTGCCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATAGCAGAAACCTGCCATGCT 1200  
QY 1201 CAGAGAACTCTAGAGATCTGAGAGATGTTCTTGGATTAACCTAAATAGCAGATTGACA 1260  
DB 1201 CAGAGAACTCTAGAGATCTGAGAGATGTTCTTGGATTAACCTAAATAGCAGATTGACA 1260  
QY 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTATGATGATGATGATGATGATGATGATGATGATG 1320  
DB 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTATGATGATGATGATGATGATGATGATGATGATG 1320  
QY 1321 GGGAGTGTGATCAATCAATGCCAAGTACGTGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 GGGAGTGTGATCAATCAATGCCAAGTACGTGATGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 AATATCTGCTGTTCTGAGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATG 1440  
DB 1381 AATATCTGCTGTTCTGAGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATG 1440  
QY 1441 TATGTAAAGTGAAGAGATTCACCTCAAAATCACTAAGAGATATATTGAAGACAAATAT 1500  
DB 1441 TATGTAAAGTGAAGAGATTCACCTCAAAATCACTAAGAGATATATTGAAGACAAATAT 1500  
QY 1501 TTGGGAAAACCTATCGGAAGAGCAAGCCTCCCACTTAAGCCATGTAAGTGAATATC 1560  
DB 1501 TTGGGAAAACCTATCGGAAGAGCAAGCCTCCCACTTAAGCCATGTAAGTGAATATC 1560  
QY 1561 TAATTTATGAGAGATTTGTTACTGAGCAGATATTAATCAAGAGGCTCCCTCACAAATA 1620  
DB 1561 TAATTTATGAGAGATTTGTTACTGAGCAGATATTAATCAAGAGGCTCCCTCACAAATA 1620  
QY 1621 AATTTAAGCCTAAAGAGAGACCTACATCAGGCTTCACTCTGAGATTTTATCAAGAAAG 1680  
DB 1621 AATTTAAGCCTAAAGAGAGACCTACATCAGGCTTCACTCTGAGATTTTATCAAGAAAG 1680  
QY 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAAGGAACTAAACAGGAGC 1740  
DB 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAAGGAACTAAACAGGAGC 1740  
QY 1741 AGATGGTCAAGTGAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
DB 1741 AGATGGTCAAGTGAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
QY 1801 CTATTTGAGATGAGAAAAATCTTAACCAATAGAACTACGCAAAAGAAATCTGCTTTCA 1860  
DB 1801 CTATTTGAGATGAGAAAAATCTTAACCAATAGAACTACGCAAAAGAAATCTGCTTTCA 1860  
QY 1861 AAAGGAAAGCTGAACTATTAAGAGAGATTAAGCAATATGGAATCGAAATTAATATCC 1920  
DB 1861 AAAGGAAAGCTGAACTATTAAGAGAGATTAAGCAATATGGAATCGAAATTAATATCC 1920  
QY 1921 ACAATTTCAAAAGACCTTAATAAAGATTAAGGCTGAGAGAAAGTCTTCTACAGGCAATATC 1980  
DB 1921 ACAATTTCAAAAGACCTTAATAAAGATTAAGGCTGAGAGAAAGTCTTCTACAGGCAATATC 1980  
QY 1981 ATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
DB 1981 ATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040

QY 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACCAACCAATGCCAGTCA 2100  
| | | | |  
Db 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACCAACCAATGCCAGTCA 2100  
QY 2101 GGCACAGCAGAAACCTACACTCATGGAAGTTAAAGAACCTGCACTGAGCCAGANAGA 2160  
| | | | |  
Db 2101 GGCACAGCAGAAACCTACACTCATGGAAGTTAAAGAACCTGCACTGAGCCAGANAGA 2160  
QY 2161 GTAAAGGCCAATGACAGACAGTAAGCAATGACAGTATCTTCCAGAGCTCA 2220  
| | | | |  
Db 2161 GTAAAGGCCAATGACAGACAGTAAGCAATGACAGTATCTTCCAGAGCTCA 2220  
QY 2221 AGTTACCAATGCACCTGTTCTTTTACTAAGTGTCAATATCCAGTGAATTAAGAAAT 2280  
| | | | |  
Db 2221 AGTTACCAATGCACCTGTTCTTTTACTAAGTGTCAATATCCAGTGAATTAAGAAAT 2280  
QY 2281 TTGTCATTCCTTAGCTTCCAGAGAGAAAGAAAGAGAACTAGAACAGTTAAAGTGT 2340  
| | | | |  
Db 2281 TTGTCATTCCTTAGCTTCCAGAGAGAAAGAAAGAGAACTAGAACAGTTAAAGTGT 2340  
QY 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTAAAGTGAAGAGAGGTTTGCAGAGTG 2400  
| | | | |  
Db 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTAAAGTGAAGAGAGGTTTGCAGAGTG 2400  
QY 2401 AAAGATCTGTAGAGAGTACAGTATTTGATGCTGTAAGTATTAAGGACACTCAG 2460  
| | | | |  
Db 2401 AAAGATCTGTAGAGAGTACAGTATTTGATGCTGTAAGTATTAAGGACACTCAG 2460  
QY 2461 AAAGATCTGCTTACTGGAAGTTAGCACTCTTAGGAGAGCAAAAACAGAACCAATTAAT 2520  
| | | | |  
Db 2461 AAAGATCTGCTTACTGGAAGTTAGCACTCTTAGGAGAGCAAAAACAGAACCAATTAAT 2520  
QY 2521 GTGTAGTCAAGTGTGACAGATTTGAAAAACCCCAAGAGGACTAATTCATGTTTCCAGAG 2580  
| | | | |  
Db 2521 GTGTAGTCAAGTGTGACAGATTTGAAAAACCCCAAGAGGACTAATTCATGTTTCCAGAG 2580  
QY 2581 ATAAATGAAGTACACAGAGGCTTTAAGTATCATCTGGACATGAAGTTAACCAAGCTC 2640  
| | | | |  
Db 2581 ATAAATGAAGTACACAGAGGCTTTAAGTATCATCTGGACATGAAGTTAACCAAGCTC 2640  
QY 2641 GGGAAACAAAGCATTAATAATGGAAGAAAGTGAAGTGTGATGTTTGCAGATTAAT 2700  
| | | | |  
Db 2641 GGGAAACAAAGCATTAATAATGGAAGAAAGTGAAGTGTGATGTTTGCAGATTAAT 2700  
QY 2701 TCAAGGTTTCAAGGGCCAGTATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
| | | | |  
Db 2701 TCAAGGTTTCAAGGGCCAGTATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
QY 2761 AATGTGCAACATTTCTGCCACTCTGGCTCTTAAGAAACCAAGTCCAAAGTCACTT 2820  
| | | | |  
Db 2761 AATGTGCAACATTTCTGCCACTCTGGCTCTTAAGAAACCAAGTCCAAAGTCACTT 2820  
QY 2821 TTGATGTGAACAAAGAGAAATCAAGAAAGAAATGAGTCTAATTCAGAGCTGTAC 2880  
| | | | |  
Db 2821 TTGATGTGAACAAAGAGAAATCAAGAAAGAAATGAGTCTAATTCAGAGCTGTAC 2880  
QY 2881 AGACAGTTAATATCACTCAGGCTTTCTGTGTTGTGTCAGAAAGATTAAGCAGTGTATA 2940  
| | | | |  
Db 2881 AGACAGTTAATATCACTCAGGCTTTCTGTGTTGTGTCAGAAAGATTAAGCAGTGTATA 2940  
QY 2941 ATGCCAAATGTATGATCAAGAGAGGCTGTAGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
| | | | |  
Db 2941 ATGCCAAATGTATGATCAAGAGAGGCTGTAGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
QY 3001 AGCAAGAGGAGCTCTTACTCTCAATTAACATGAGCTTTTACAAAACCCATATCTGTATAC 3060  
| | | | |  
Db 3001 AGCAAGAGGAGCTCTTACTCTCAATTAACATGAGCTTTTACAAAACCCATATCTGTATAC 3060  
QY 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACAAATGTAAGAAATATGCTGAGAGG 3120  
| | | | |  
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACAAATGTAAGAAATATGCTGAGAGG 3120  
QY 3121 AAAACTTTGAGAACATTCATGTCACCTGAAAGAAATGGAATGAGAACATTCGCA 3180  
| | | | |

Db 3121 AAAACTTTGAGAACATTCATGTCACCTGAAAGAAATGGAATGAGAACATTCGCA 3180  
| | | | |  
QY 3181 GTACAGTGAACCAATTAAGCCGTAAATTAACATTAAGAAATATTTTAAAGAGCCAGCT 3240  
| | | | |  
Db 3181 GTACAGTGAACCAATTAAGCCGTAAATTAACATTAAGAAATATTTTAAAGAGCCAGCT 3240  
QY 3241 CAAGCAATTAATTAAGTACAGTTCAGTACATTAAGAGTGGGCCGATTAATTAATGAA 3300  
| | | | |  
Db 3241 CAAGCAATTAATTAAGTACAGTTCAGTACATTAAGAGTGGGCCGATTAATTAATGAA 3300  
QY 3301 TAGGTTCCAGTATGATTAAGCAATTAAGCAAGAACTAGTGAAGAAACAGAGGCCAAATTTGA 3360  
| | | | |  
Db 3301 TAGGTTCCAGTATGATTAAGCAATTAAGCAAGAACTAGTGAAGAAACAGAGGCCAAATTTGA 3360  
QY 3361 ATGCTATGCTTAGATTAAGGAGTTTTCGCAACCTGAGGCTATTAACAAAGCTTCTCGAA 3420  
| | | | |  
Db 3361 ATGCTATGCTTAGATTAAGGAGTTTTCGCAACCTGAGGCTATTAACAAAGCTTCTCGAA 3420  
QY 3421 GTAATTTGTAAGCATCTGTAATTAAGCAAGAAATGAAGAGTACGTTGAGACTGTTA 3480  
| | | | |  
Db 3421 GTAATTTGTAAGCATCTGTAATTAAGCAAGAAATGAAGAGTACGTTGAGACTGTTA 3480  
QY 3481 ATACAGATTTCTCTCCATATCTGATTTAGATTAAGTGAAGAGGCTATGGAAGTACTC 3540  
| | | | |  
Db 3481 ATACAGATTTCTCTCCATATCTGATTTAGATTAAGTGAAGAGGCTATGGAAGTACTC 3540  
QY 3541 ATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATTAAGG 3600  
| | | | |  
Db 3541 ATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATTAAGG 3600  
QY 3601 AAGATACGATTTTCTGTAATGATTAAGGAAGTTCGCTTTTGAACAAAGC 3660  
| | | | |  
Db 3601 AAGATACGATTTTCTGTAATGATTAAGGAAGTTCGCTTTTGAACAAAGC 3660  
QY 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGGCTTTACACCATACACATTTGGCTCAGG 3720  
| | | | |  
Db 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGGCTTTACACCATACACATTTGGCTCAGG 3720  
QY 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTGCTGAAAGAGACATTAATCTAGTGAAGT 3780  
| | | | |  
Db 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTGCTGAAAGAGACATTAATCTAGTGAAGT 3780  
QY 3781 AAGACCTCCCTGCTTCCACACCTGTTATTTGTAAGTAATTAACATTAATCTGCTGCT 3840  
| | | | |  
Db 3781 AAGACCTCCCTGCTTCCACACCTGTTATTTGTAAGTAATTAACATTAATCTGCTGCT 3840  
QY 3841 CTACTAGGCATAGCACCGTGTCTACCGAGTGTCTGTAAGAAACACAGAGAGAAATTTAT 3900  
| | | | |  
Db 3841 CTACTAGGCATAGCACCGTGTGTACCGAGTGTCTGTAAGAAACACAGAGAGAAATTTAT 3900  
QY 3901 TATCATTTGAAGAAATAGCTTAATGATGCTGACAGTAAATTAATGCAAGAGCATCTC 3960  
| | | | |  
Db 3901 TATCATTTGAAGAAATAGCTTAATGATGCTGACAGTAAATTAATGCAAGAGCATCTC 3960  
QY 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTGCTGAGTGTGTTTCTTCCAGAGTCA 4020  
| | | | |  
Db 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTGCTGAGTGTGTTTCTTCCAGAGTCA 4020  
QY 4021 GTGAATTTGAAGACTTTGACGCAAAATTAACACCCAGAGATCTTTTGAATGTTCTT 4080  
| | | | |  
Db 4021 GTGAATTTGAAGACTTTGACGCAAAATTAACACCCAGAGATCTTTTGAATGTTCTT 4080  
QY 4081 CCAAAACAAATAGGATAGTCTGAAAGCCAGAGGAGTTGCTGAGTGAAGAAATTTG 4140  
| | | | |  
Db 4081 CCAAAACAAATAGGATAGTCTGAAAGCCAGAGGAGTTGCTGAGTGAAGAAATTTG 4140  
QY 4141 TTTGATGATGATGAAGAAAGAGAGGCTTGAAGAAATTAATCAAGAGCAAGCA 4200  
| | | | |  
Db 4141 TTTGATGATGATGAAGAAAGAGAGGCTTGAAGAAATTAATCAAGAGCAAGCA 4200  
QY 4201 TGAATTTAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAGAACGCTCTCTGAG 4260  
| | | | |

```

Db 4201 TGAATCAAACTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAGCCGTCTGTGAG 4260
OY 4261 ACTGTCAGGGGCTATCTCTCAGAGTGTACATTTTAAACCACCTCAGCAGAGGGATACCATGC 4320
Db 4261 ACTGTCAGGGGCTATCTCTCAGAGTGTACATTTTAAACCACCTCAGCAGAGGGATACCATGC 4320
OY 4321 AACATTAACCTGATTAAGGCTCCAGCAGAGAAATGGCTGAACCTGTGTTAGAAGC 4380
Db 4321 AACATTAACCTGATTAAGGCTCCAGCAGAGAAATGGCTGAACCTGTGTTAGAAGC 4380
OY 4381 ATGGGAGCCAGCCTTTCTACAGCTAACCTTGCATATATAGTGACCTTCTGCCCTTGAG 4440
Db 4381 ATGGGAGCCAGCCTTTCTACAGCTAACCTTGCATATATAGTGACCTTCTGCCCTTGAG 4440
OY 4441 ACCGCGCAAAATCCAGAAACAAAGCAGATCAGAAAAAGAGATTAATCTTACAGAAAGTA 4500
Db 4441 ACCGCGCAAAATCCAGAAACAAAGCAGATCAGAAAAAGAGATTAATCTTACAGAAAGTA 4500
OY 4501 GTGAATACCTATTAAGCCAGAAATCCAGAGGCTTTTGTGTGACAGTTTGAAGTGTCTG 4560
Db 4501 GTGAATACCTATTAAGCCAGAAATCCAGAGGCTTTTGTGTGACAGTTTGAAGTGTCTG 4560
OY 4561 CAGATAGTCTACAGTAAAAATAAGAACAGGAGGAGAAAGGTCATCCCTTCTTAAT 4620
Db 4561 CAGATAGTCTACAGTAAAAATAAGAACAGGAGGAGAAAGGTCATCCCTTCTTAAT 4620
OY 4621 GCCCATCATTTAGATAGTGTGTATGATGACAGAGTGTGTGAGAGGAGCAACAGCTGAG 4680
Db 4621 GCCCATCATTTAGATAGTGTGTATGATGACAGAGTGTGTGAGAGGAGCAACAGCTGAG 4680
OY 4681 ACTTACCATCTCAAGAGAGGCTCATTAAGTGTGTGATGTGAGAGGAGCAACAGCTGAG 4740
Db 4681 ACTTACCATCTCAAGAGAGGCTCATTAAGTGTGTGATGTGAGAGGAGCAACAGCTGAG 4740
OY 4741 AGTGTGGGCCACAGATTTGAGGAGAAACATCTTCTTCCAGAGGAGCAACATCTAGAGG 4800
Db 4741 AGTGTGGGCCACAGATTTGAGGAGAAACATCTTCTTCCAGAGGAGCAACATCTAGAGG 4800
OY 4801 CCCCTTACCTGGAATCTGGAATCAAGCCCTTCTCTGATGACCCCTGAATCTGATCTTGTG 4860
Db 4801 CCCCTTACCTGGAATCTGGAATCAAGCCCTTCTCTGATGACCCCTGAATCTGATCTTGTG 4860
OY 4861 AAGCAGAGCCGCCAGAGTCAAGCTGTGTGCAACATACCATCTTCAACCTCTGCAATGA 4920
Db 4861 AAGCAGAGCCGCCAGAGTCAAGCTGTGTGCAACATACCATCTTCAACCTCTGCAATGA 4920
OY 4921 AAGTTCGCCAATTAAGAGTTGCAAGTCTGCCAGAGTCCAGCTGCTCTCATATCTATCG 4980
Db 4921 AAGTTCGCCAATTAAGAGTTGCAAGTCTGCCAGAGTCCAGCTGCTCTCATATCTATCG 4980
OY 4981 ATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAGCCAGAAATTTGAC 5040
Db 4981 ATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAGCCAGAAATTTGAC 5040
OY 5041 CTTCACAGAAAGGGTCAACAAAGAAATGTCCATGTGTGTGTGCTGAGCCCGAGAG 5100
Db 5041 CTTCACAGAAAGGGTCAACAAAGAAATGTCCATGTGTGTGTGCTGAGCCCGAGAG 5100
OY 5101 AATTATGCTGCTGTACAAAGTTTGCAGAAACACACATCATTAACTTAATCTAATTA 5160
Db 5101 AATTATGCTGCTGTGTACAAAGTTTGCAGAAACACACATCATTAACTTAATCTAATTA 5160
OY 5161 CTGAAGAGACTACTCATGTGTGTATGAAGACAGATGCTGAGTTGTGTGTGAAGGAGAC 5220
Db 5161 CTGAAGAGACTACTCATGTGTGTATGAAGACAGATGCTGAGTTGTGTGTGAAGGAGAC 5220
OY 5221 TGAATATTTTCTAGGAATTCGGGAGAGAAATGGGTAGTTAGCTATTTCTGGGTGACCC 5280
Db 5221 TGAATATTTTCTAGGAATTCGGGAGAGAAATGGGTAGTTAGCTATTTCTGGGTGACCC 5280
OY 5281 AGTCTATTAAGAAAGAAATGCTGAATGAGCATGATTGGAAGTCAAGAGAGATGTGG 5340
Db 5281 AGTCTATTAAGAAAGAAATGCTGAATGAGCATGATTGGAAGTCAAGAGAGATGTGG 5340

```

```

OY 5341 TCAATGGAAGAACACCAGAGTCCAAAGCAGCAGAGAGAAATCCAGAGAGAAAGATCT 5400
Db 5341 TCAATGGAAGAAACACCAGAGTCCAAAGCAGCAGAGAGAAATCCAGAGAGAAAGATCT 5400
OY 5401 TCAGGGGCTAGAAATCTGTTCATAGGAGCCCTTCCACCAACATGCCAGATCAACTG 5460
Db 5401 TCAGGGGCTAGAAATCTGTTCATAGGAGCCCTTCCACCAACATGCCAGATCAACTG 5460
OY 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5520
Db 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5520
OY 5521 GCACAGTGTCCACCAATTTGTTGTGTCAGCAGATGCTTGACAGAGAGAAATGGCT 5580
Db 5521 GCACAGTGTCCACCAATTTGTTGTGTCAGCAGATGCTTGACAGAGAGAAATGGCT 5580
OY 5581 TCATATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGTGTGACA 5640
Db 5581 TCATATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGTGTGACA 5640
OY 5641 GTGTAGCACTACAGAGTGGCAGAGTGGCAGACCTACCTGATACCTCCAGATCCCCACA 5700
Db 5641 GTGTAGCACTACAGAGTGGCAGAGTGGCAGACCTACCTGATACCTCCAGATCCCCACA 5700
OY 5701 GCCACTACTGA 5711
Db 5701 GCCACTACTGA 5711

RESULT 2
AR112810
LOCUS AR112810 5711 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 5 from patent US 6130322.
ACCESSION AR112810
KEYWORDS AR112810.1 GI:14092710
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5711)
AUTHORS Murphy, P.D., Allen, A.C.P., Alvares, C.P., Critz, B.S., Olson, S.J.,
Thurber, D., and Zeng, B.
TITLE Coding sequences of the human BRCA1 gene
JOURNAL Patent: US 6130322-A 5 10-Oct-2000;
FEATURES
source 1. 5711
BASE COUNT 1956 a 1098 c 1274 g 1383 t
ORIGIN

Query Match 100.0%; Score 5709.4; DB 6; Length 5711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCTGCTGAGAGCTCTCTGACCCCGACAGGCTGTGGGGTTTCTCAGATACTGGGCC 60
Db 1 AGCTGCTGAGAGCTCTCTGACCCCGACAGGCTGTGGGGTTTCTCAGATACTGGGCC 60
OY 61 CCTGCGCTCAGAGAGGCTTCACTCTGTCTGTGTGGGTAAAGTTCAATTTGGAACAGAAAGAA 120
Db 61 CCTGCGCTCAGAGAGGCTTCACTCTGTCTGTGTGGGTAAAGTTCAATTTGGAACAGAAAGAA 120
OY 121 TGAATTTATCTGCTCTTCTGCGCTTGAAGAAGTCAAAATGTCATTAATGCTATGCGAGAAA 180
Db 121 TGAATTTATCTGCTCTTCTGCGCTTGAAGAAGTCAAAATGTCATTAATGCTATGCGAGAAA 180
OY 181 TCTTAGAGTGTCCCATCTGTCTGTGAGTGTGATCAAGAACTGTCTCCAAAGTGTGACC 240
Db 181 TCTTAGAGTGTCCCATCTGTCTGTGAGTGTGATCAAGAACTGTCTCCAAAGTGTGACC 240
OY 241 ACATATTTTGAATTTTGCATGCTGAAGTCTTCAACAGAGAAAGGCGCTTCAAGT 300
Db 241 ACATATTTTGAATTTTGCATGCTGAAGTCTTCAACAGAGAAAGGCGCTTCAAGT 300

```



Db 241 ACATTTTTCGAATTTTGCATGCTGAAACTTCTCACACAGAGAAAGGCCCTTCACAGT 300  
Qy 301 GTCCCTTATGTAGATATATTAACCAAGAGGCTTACAGAAAGTACGAGATTAGTC 360  
Db 301 GTCCCTTATGTAGATATATTAACCAAGAGGCTTACAGAAAGTACGAGATTAGTC 360  
Qy 361 AACTGTTGAAGAGCTATTGAAATCAATTGCTTTGCTTACGCTGACACAGGTTGGAGT 420  
Db 361 AACTGTTGAAGAGCTATTGAAATCAATTGCTTTGCTTACGCTGACACAGGTTGGAGT 420  
Qy 421 ATGCAACAGCTATATTTTGCAGAAAGAAATTAAGTCTCCGTAACATCTAAAGATG 480  
Db 421 ATGCAACAGCTATATTTTGCAGAAAGAAATTAAGTCTCCGTAACATCTAAAGATG 480  
Qy 481 AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCAGAAAGACTTCTACAGATG 540  
Db 481 AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCAGAAAGACTTCTACAGATG 540  
Qy 541 AACCCGAAATCTCTCTGAGAGAAACCACTCTAGTGTCCAACTCTCTAACTTGGAA 600  
Db 541 AACCCGAAATCTCTCTGAGAGAAACCACTCTAGTGTCCAACTCTCTAACTTGGAA 600  
Qy 601 CTGTGAGACCTCTAGAGCAAGACGCGGATACAACTCCAAAGACGCTCTCTACATG 660  
Db 601 CTGTGAGACCTCTAGAGCAAGACGCGGATACAACTCCAAAGACGCTCTCTACATG 660  
Qy 661 AATTGGATCTGATCTCTCTAGAGTACCGTTAATAGCAACTATTGCGAGTGGAG 720  
Db 661 AATTGGATCTGATCTCTCTAGAGTACCGTTAATAGCAACTATTGCGAGTGGAG 720  
Qy 721 ATCAAGAAATGTTCAAATCAACCCCTCAAGAGAACCGAGGATGAATTCAGTTTGTG 780  
Db 721 ATCAAGAAATGTTCAAATCAACCCCTCAAGAGAACCGAGGATGAATTCAGTTTGTG 780  
Qy 781 CAAAAAGGCTGTGTAATTTTCTGAGAGGATGTAACAATCTAATCATCTAC 840  
Db 781 CAAAAAGGCTGTGTAATTTTCTGAGAGGATGTAACAATCTAATCATCTAC 840  
Qy 841 CCAGTAATATGATTTGAAACACCACTGAGAGCGTGCAGTGCAGAGGCTCAGAAAGT 900  
Db 841 CCAGTAATATGATTTGAAACACCACTGAGAGCGTGCAGTGCAGAGGCTCAGAAAGT 900  
Qy 901 ATCAGGATGTTCTGTTTCAAACTGTCATGTGAGACCATGTGGCACAATCTCATGCCA 960  
Db 901 ATCAGGATGTTCTGTTTCAAACTGTCATGTGAGACCATGTGGCACAATCTCATGCCA 960  
Qy 961 GCTATTTACAGCATGAGACAGCACTTATCTACTCAATAGAGCAAGATGATAGAA 1020  
Db 961 GCTATTTACAGCATGAGACAGCACTTATCTACTCAATAGAGCAAGATGATAGAA 1020  
Qy 1021 AGGCTGATTTCTGTATTAAGCAAAAGCAGCTGCTTACAGAGGACCAATACATGAT 1080  
Db 1021 AGGCTGATTTCTGTATTAAGCAAAAGCAGCTGCTTACAGAGGACCAATACATGAT 1080  
Qy 1081 GGGCTGGAAGTAAAGAAATGATAGGCGGACTCCAGCAGAAAAAAGGTAG 1140  
Db 1081 GGGCTGGAAGTAAAGAAATGATAGGCGGACTCCAGCAGAAAAAAGGTAG 1140  
Qy 1141 ATCTGAAGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGAGAAACTGCCATCT 1200  
Db 1141 ATCTGAAGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGAGAAACTGCCATCT 1200  
Qy 1201 CAGAGAACTCTAGAGATCTAGAGATGTTGGATACACTAAATAGCAGCATTCGA 1260  
Db 1201 CAGAGAACTCTAGAGATCTAGAGATGTTGGATACACTAAATAGCAGCATTCGA 1260  
Qy 1261 AAGTTAATGAGTGTTCAGAGAGTGAATGAGTTCTGTGATGACCTCAGATGATG 1320  
Db 1261 AAGTTAATGAGTGTTCAGAGAGTGAATGAGTTCTGTGATGACCTCAGATGATG 1320  
Qy 1321 GGGAGTCTGATCAAAATGCCAAAGTACGATGATTTGAGAGGCTTCTAAATAGGATG 1380  
Db 1321 GGGAGTCTGATCAAAATGCCAAAGTACGATGATTTGAGAGGCTTCTAAATAGGATG 1380

Qy 1381 AATATTCGTCTTCAAGAGAAATAGACTTACTGGCCAGTGCATCTCATGAGGCTTAA 1440  
Db 1381 AATATTCGTCTTCAAGAGAAATAGACTTACTGGCCAGTGCATCTCATGAGGCTTAA 1440  
Qy 1441 TATGTAAAGTGAAGAGTCTCAATATCAGTGAAGATATATTTGAGCAAAATAT 1500  
Db 1441 TATGTAAAGTGAAGAGTCTCAATATCAGTGAAGATATATTTGAGCAAAATAT 1500  
Qy 1501 TTGGGAAACCTATGGAAGAGGACGCTCCCACTTAAGCATGAACTGAAATC 1560  
Db 1501 TTGGGAAACCTATGGAAGAGGACGCTCCCACTTAAGCATGAACTGAAATC 1560  
Qy 1561 TAATATATGAGCATTTGTTACTGAGCCAGATATATCAAGAGGCTCCCTCAGAAATA 1620  
Db 1561 TAATATATGAGCATTTGTTACTGAGCCAGATATATCAAGAGGCTCCCTCAGAAATA 1620  
Qy 1621 AATTAAAGCGTAAAGAGACCTACATCAGGCTTCATCTGAGGATTTTATCAAGAAAG 1680  
Db 1621 AATTAAAGCGTAAAGAGACCTACATCAGGCTTCATCTGAGGATTTTATCAAGAAAG 1680  
Qy 1681 CAGATTTGGAGTCAAAAGACTCCTGAAATGATTAATCAGGAACTAAACCAAGGAGC 1740  
Db 1681 CAGATTTGGAGTCAAAAGACTCCTGAAATGATTAATCAGGAACTAAACCAAGGAGC 1740  
Qy 1741 AGAATGCTCAAGTGAATATTTACTAATAGTGCATAGAGATTAACCAAGGATGAT 1800  
Db 1741 AGAATGCTCAAGTGAATATTTACTAATAGTGCATAGAGATTAACCAAGGATGAT 1800  
Qy 1801 CTATTCAGATGAGAAAAATCCTAACCATAATGAATCACTGCAAAAGATGTCTTCA 1860  
Db 1801 CTATTCAGATGAGAAAAATCCTAACCATAATGAATCACTGCAAAAGATGTCTTCA 1860  
Qy 1861 AAACGAAGCTGAACCTTAAGCAGAGATTAAGCAATGGAACCTGATTAATATCC 1920  
Db 1861 AAACGAAGCTGAACCTTAAGCAGAGATTAAGCAATGGAACCTGATTAATATCC 1920  
Qy 1921 ACAATTCAAAAGCACCTTAAAGAGTGAAGGAGGAGTCTTCAACAGGATATTC 1980  
Db 1921 ACAATTCAAAAGCACCTTAAAGAGTGAAGGAGGAGTCTTCAACAGGATATTC 1980  
Qy 1981 ATGCGCTTGAAGTACTAGTACAGTGAAGATCTTAAGCCACCTTAATGTACTGATTC 2040  
Db 1981 ATGCGCTTGAAGTACTAGTACAGTGAAGATCTTAAGCCACCTTAATGTACTGATTC 2040  
Qy 2041 TTGATAGTGTCTTCAAGTGAAGATTAAGAAAAAGTAAACCAAGATGCAAGTCA 2100  
Db 2041 TTGATAGTGTCTTCAAGTGAAGATTAAGAAAAAGTAAACCAAGATGCAAGTCA 2100  
Qy 2101 GGCACAGCAGAAACCTTACATGATGAAGTGAAGAACTTCACTGAGCAGAGAGA 2160  
Db 2101 GGCACAGCAGAAACCTTACATGATGAAGTGAAGAACTTCACTGAGCAGAGAGA 2160  
Qy 2161 GTACACACCAATGAGACAGACAGTAAAGACATGACAGTGTACTTTCCAGAGCTGA 2220  
Db 2161 GTACACACCAATGAGACAGACAGTAAAGACATGACAGTGTACTTTCCAGAGCTGA 2220  
Qy 2221 ACTTAACAATGACACTGTTCTTTTACTAATGTGTTCAATATACGATGATTAAGAT 2280  
Db 2221 ACTTAACAATGACACTGTTCTTTTACTAATGTGTTCAATATACGATGATTAAGAT 2280  
Qy 2281 TTGTCATCTCAAGCTTCCAGAGAGAGAAAAAGAGAAACTTAAGATGAT 2340  
Db 2281 TTGTCATCTCAAGCTTCCAGAGAGAGAAAAAGAGAAACTTAAGATGAT 2340  
Qy 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGCTTTTGCAGATG 2400  
Db 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGCTTTTGCAGATG 2400  
Qy 2401 AAAGATCTGATGAGTGAAGTATTTATGATGATGATGATGATGATGATGATGATG 2460  
Db 2401 AAAGATCTGATGAGTGAAGTATTTATGATGATGATGATGATGATGATGATGATG 2460

```
QY 2461 AAGATATCTGTTACTGGAAGTTAGCACTCTAGGGAGGCAAAAAAGACAAATAAAT 2520
    |||||
Db 2461 AAGATATCTGTTACTGGAAGTTAGCACTCTAGGGAGGCAAAAAAGACAAATAAAT 2520
QY 2521 GTGTGACTGAGTGGGACATTTGAAAAACCCAAAGGCACTAATTCATGTTGTTCCAAAG 2580
    |||||
Db 2521 GTGTGACTGAGTGGGACATTTGAAAAACCCAAAGGCACTAATTCATGTTGTTCCAAAG 2580
QY 2581 ATATATGAATATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAGTTAAACCACTG 2640
    |||||
Db 2581 ATATATGAATATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAGTTAAACCACTG 2640
QY 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGCTCACTATTTGCGAATACAT 2700
    |||||
Db 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGCTCACTATTTGCGAATACAT 2700
QY 2701 TCAGGTTTCAAGGCGCACTATTTGCTGTTTCAAAATCCGGAATTCAGAAAGAG 2760
    |||||
Db 2701 TCAGGTTTCAAGGCGCACTATTTGCTGTTTCAAAATCCGGAATTCAGAAAGAG 2760
QY 2761 AATGTGCAACATTTCTGCGCACCTGGGTCCTTAAAGAAACAAAGTCCAAAGTCACTT 2820
    |||||
Db 2761 AATGTGCAACATTTCTGCGCACCTGGGTCCTTAAAGAAACAAAGTCCAAAGTCACTT 2820
QY 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCCTGTAC 2880
    |||||
Db 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCCTGTAC 2880
QY 2881 AGACATTTATATCACTGCGAGGCTTCCCTGGTGGTCAAGAAAGATTAAGCCATTTGATA 2940
    |||||
Db 2881 AGACATTTATATCACTGCGAGGCTTCCCTGGTGGTCAAGAAAGATTAAGCCATTTGATA 2940
QY 2941 ATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCACTGTTAGAGCA 3000
    |||||
Db 2941 ATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCACTGTTAGAGCA 3000
QY 3001 AGCAAACTGACATCTACTCCAAATTAACATGACCTTTTCAAAACCCATATGCTATAC 3060
    |||||
Db 3001 AGCAAACTGACATCTACTCCAAATTAACATGACCTTTTCAAAACCCATATGCTATAC 3060
QY 3061 CACCACTTTTCCATCAAGTCAATTTGTTTAAACTAATTAAGAAATAATCTGCTAAGG 3120
    |||||
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTTAAACTAATTAAGAAATAATCTGCTAAGG 3120
QY 3061 CACCACTTTTCCATCAAGTCAATTTGTTTAAACTAATTAAGAAATAATCTGCTAAGG 3120
    |||||
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTTAAACTAATTAAGAAATAATCTGCTAAGG 3120
QY 3121 AAAAATTTGAGAAACATTCATGTCACCTGAAAAGAAATGGAAATGAGAACATTTCAA 3180
    |||||
Db 3121 AAAAATTTGAGAAACATTCATGTCACCTGAAAAGAAATGGAAATGAGAACATTTCAA 3180
QY 3181 GTACAGTAGACCAATTTAGCCGTAAATACATTTAGAGAAATGTTTTTAAAGAGCCAGCT 3240
    |||||
Db 3181 GTACAGTAGACCAATTTAGCCGTAAATACATTTAGAGAAATGTTTTTAAAGAGCCAGCT 3240
QY 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAAGTAA 3300
    |||||
Db 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAAGTAA 3300
QY 3301 TAGGTTCCAGTGTGAAGAAACATTCAGACAGACATAGTAGAAACAGAGGCGCAAAATGA 3360
    |||||
Db 3301 TAGGTTCCAGTGTGAAGAAACATTCAGACAGACATAGTAGAAACAGAGGCGCAAAATGA 3360
QY 3361 ATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTCTAATAACAAAGCTTCTCTGAA 3420
    |||||
Db 3361 ATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTCTAATAACAAAGCTTCTCTGAA 3420
QY 3421 GTAATTTGAAGCATCTGGAATTAAGAAAGCAAGAAATATGAAGTAGTTCAGACCTGTA 3480
    |||||
Db 3421 GTAATTTGAAGCATCTGGAATTAAGAAAGCAAGAAATATGAAGTAGTTCAGACCTGTA 3480
QY 3481 ATACAGATTTCTCTCATATCTGATTTAGATTAAGTAAAGAGCCATAGGGAAGTAGTAC 3540
    |||||
Db 3481 ATACAGATTTCTCTCATATCTGATTTAGATTAAGTAAAGAGCCATAGGGAAGTAGTAC 3540
QY 3541 ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAGG 3600
    |||||
```



Db 661 AATTGGATCTGATCTTCTGAGATACCGTTAATAGGCACTTATTGCAGTGTGGAG 720  
QY 721 ATCAGAAATTGTTACAAATGACCCCTCAAGAACACAGGATGAATACGTTTGGATTCTG 780  
Db 721 ATCAGAAATTGTTACAAATGACCCCTCAAGAACACAGGATGAATACGTTTGGATTCTG 780  
QY 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGACGATGTAAACAAATACTGAACATCATCAAC 840  
Db 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGACGATGTAAACAAATACTGAACATCATCAAC 840  
QY 841 CCAGTAATATGATTTGAACACCCACGAGAAAGCGCTGACGTGAGAGCATCCAAAAAGT 900  
Db 841 CCAGTAATATGATTTGAACACCCACGAGAAAGCGCTGACGTGAGAGCATCCAAAAAGT 900  
QY 901 ATCAGGATGCTTGTCTTCAAACTTGATGTGAGCCATGTGGCACAATATCATGCA 960  
Db 901 ATCAGGATGCTTGTCTTCAAACTTGATGTGAGCCATGTGGCACAATATCATGCA 960  
QY 961 GCTCATTTACAGCATGAGACAGCAGTTTACTCACTAAAGACAGAAATGATAGAAA 1020  
Db 961 GCTCATTTACAGCATGAGACAGCAGTTTACTCACTAAAGACAGAAATGATAGAAA 1020  
QY 1021 AGGCTGAATCTGTAAATTAAGCAACAGCGCTGGCTTAGCAAGGACATCAACAGAT 1080  
Db 1021 AGGCTGAATCTGTAAATTAAGCAACAGCGCTGGCTTAGCAAGGACATCAACAGAT 1080  
QY 1081 GGGCTGGAATGAGAAACATGTATGATAGGGGACCTCCAGCAGCAAGAAAAAGGTAG 1140  
Db 1081 GGGCTGGAATGAGAAACATGTATGATAGGGGACCTCCAGCAGCAAGAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTCTGTGAGAGAAAGATGGAATAGCAGAAAAGTCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTCTGTGAGAGAAAGATGGAATAGCAGAAAAGTCCATGCT 1200  
QY 1201 CAGAGAAATCCTAGATACTGAGATGTTCTTGATAGACATTAATAGAGCATTTAGA 1260  
Db 1201 CAGAGAAATCCTAGATACTGAGATGTTCTTGATAGACATTAATAGAGCATTTAGA 1260  
QY 1261 AAGTTAATGATGTTTCCAGAAAGTATGATAGCTGTTAGGTTCTGATGATCAATGATG 1320  
Db 1261 AAGTTAATGATGTTTCCAGAAAGTATGATAGCTGTTAGGTTCTGATGATCAATGATG 1320  
QY 1321 GGGAGCTGAATCAATGCGCAAGTATGATGATGATGATGATGATGATGATGATGATG 1380  
Db 1321 GGGAGCTGAATCAATGCGCAAGTATGATGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 AATATCTGCTTCTCAGAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 1440  
Db 1381 AATATCTGCTTCTCAGAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 1440  
QY 1441 TATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGCAAAATAT 1500  
Db 1441 TATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGCAAAATAT 1500  
QY 1501 TTGGGAAAACTATCGGAAGAGGCAAGCTCCCAACTTAAGCCATGTAAGTGAATATC 1560  
Db 1501 TTGGGAAAACTATCGGAAGAGGCAAGCTCCCAACTTAAGCCATGTAAGTGAATATC 1560  
QY 1561 TAATTATAGACATTTGTTACTGAGCAACAGATTAATACAGAGCGTCCCTCACAATA 1620  
Db 1561 TAATTATAGACATTTGTTACTGAGCAACAGATTAATACAGAGCGTCCCTCACAATA 1620  
QY 1621 AATTAAGGCTAAAGAGAGCTACATCAGGCTTCATCCTGAGATTTTATCAAAAAG 1680  
Db 1621 AATTAAGGCTAAAGAGAGCTACATCAGGCTTCATCCTGAGATTTTATCAAAAAG 1680  
QY 1681 CAGATTTGGCACTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTAACCAAGCAGAG 1740  
Db 1681 CAGATTTGGCACTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTAACCAAGCAGAG 1740  
QY 1741 AGAATGTCAGTGAATATTTACTAATAGTGTGATGATGATGATGATGATGATGATGAT 1800  
Db 1741 AGAATGTCAGTGAATATTTACTAATAGTGTGATGATGATGATGATGATGATGATGAT 1800

QY 1801 CTATTTAGAAATGAGAAAAATCCTTAACCCAPAGATATCATCGAAAAAGAAATCGTTTCA 1860  
Db 1801 CTATTTAGAAATGAGAAAAATCCTTAACCCAPAGATATCATCGAAAAAGAAATCGTTTCA 1860  
QY 1861 AAAGAAAGCTGAACCTTAAGAGCAGTATTAAGCATATGAACTGGAATTAATATATCC 1920  
Db 1861 AAAGAAAGCTGAACCTTAAGAGCAGTATTAAGCATATGAACTGGAATTAATATATCC 1920  
QY 1921 ACAATTTAAAGACCTTAAGAAATAGGCTGAGAGAAAGTCTTACCAGCATATTC 1980  
Db 1921 ACAATTTAAAGACCTTAAGAAATAGGCTGAGAGAAAGTCTTACCAGCATATTC 1980  
QY 1981 ATGGCTTGAATAGTGTAGTGAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2040  
Db 1981 ATGGCTTGAATAGTGTAGTGAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2040  
QY 2041 TTGATAGTGTGTAGCAGTGAAGAGATTAAGAAAAAGTAAAGTAAAGTAAAGTAAAG 2100  
Db 2041 TTGATAGTGTGTAGCAGTGAAGAGATTAAGAAAAAGTAAAGTAAAGTAAAGTAAAG 2100  
QY 2101 GGCACAGCAAAACCTTAACCTATGAGAGGTAAAGACCTGCACTGAGCCAGAGAGA 2160  
Db 2101 GGCACAGCAAAACCTTAACCTATGAGAGGTAAAGACCTGCACTGAGCCAGAGAGA 2160  
QY 2161 GTAAACAGCCAAATGAACAGACATGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2220  
Db 2161 GTAAACAGCCAAATGAACAGACATGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2220  
QY 2221 AGTTAACAATGACCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2280  
Db 2221 AGTTAACAATGACCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2280  
QY 2281 TTGTCAATCTTACCTTCCAGAGAGAAAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2340  
Db 2281 TTGTCAATCTTACCTTCCAGAGAGAAAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2340  
QY 2341 CTATTAATGCTGAAGACCCCAAGATCTATGTTAAAGTGAAGAGGTTTGCACAACTG 2400  
Db 2341 CTATTAATGCTGAAGACCCCAAGATCTATGTTAAAGTGAAGAGGTTTGCACAACTG 2400  
QY 2401 AAGATCTGTAGAGTACAGATTAATGATGATGATGATGATGATGATGATGATGATG 2460  
Db 2401 AAGATCTGTAGAGTACAGATTAATGATGATGATGATGATGATGATGATGATGATG 2460  
QY 2461 AAGATCTGTAGAGTACAGATTAATGATGATGATGATGATGATGATGATGATGATG 2520  
Db 2461 AAGATCTGTAGAGTACAGATTAATGATGATGATGATGATGATGATGATGATGATG 2520  
QY 2521 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
Db 2521 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
QY 2581 ATATATGAATGACACAGAAAGCTTAATGATGATGATGATGATGATGATGATGATGAT 2640  
Db 2581 ATATATGAATGACACAGAAAGCTTAATGATGATGATGATGATGATGATGATGATGAT 2640  
QY 2641 GGGAAACAGCATGAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGT 2700  
Db 2641 GGGAAACAGCATGAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGT 2700  
QY 2701 TCAAGTTTCAAGAGCCGCTGATTTGCTTTTCAAAATCGAGAAATGCAAGAGG 2760  
Db 2701 TCAAGTTTCAAGAGCCGCTGATTTGCTTTTCAAAATCGAGAAATGCAAGAGG 2760  
QY 2761 AATGTGCAACATTTCTGCCCAGTCTGGTCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
Db 2761 AATGTGCAACATTTCTGCCCAGTCTGGTCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
QY 2821 TTGAATGTGAACAAAGAGAAATTAAGAAAGATGATCTTAATATCAAGCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAGAAATTAAGAAAGATGATCTTAATATCAAGCTGTAC 2880

2881 AGACAGTAAATATACACTGAGGCTTCTGCTGGTGGTCAGAAAGATAAGCCAGTTGATA 2940  
|||||  
2881 AGACAGTAAATATACACTGAGGCTTCTGCTGGTGGTCAGAAAGATAAGCCAGTTGATA 2940  
|||||  
2941 ATGCCAAATGTATATCAAGAGAGGCTAGTGTGTCATCATCTCAGTTCAGAGCA 3000  
|||||  
2941 ATGCCAAATGTATATCAAGAGAGGCTAGTGTGTCATCATCTCAGTTCAGAGCA 3000  
|||||  
3001 AGCAACAGGAGCTCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATGCTATAC 3060  
|||||  
3001 AGCAACAGGAGCTCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATGCTATAC 3060  
|||||  
3061 CACCATTCTTCCATCAAGTCATTTGTTAAACTAAATGTAGAAAATCGCTAGAGG 3120  
|||||  
3061 CACCATTCTTCCATCAAGTCATTTGTTAAACTAAATGTAGAAAATCGCTAGAGG 3120  
|||||  
3121 AAACTTGGAGACATTCATGTCACCTGAAAGAAATGGAAATGAGAACATTTCCAA 3180  
|||||  
3121 AAACTTGGAGACATTCATGTCACCTGAAAGAAATGGAAATGAGAACATTTCCAA 3180  
|||||  
3181 GTACAGTACGACATTTAGCCGTATATACATTTAGAGAAATGTTTTTAAAGAGCCAGCT 3240  
|||||  
3181 GTACAGTACGACATTTAGCCGTATATACATTTAGAGAAATGTTTTTAAAGAGCCAGCT 3240  
|||||  
3241 CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCAGTATTAATGAAA 3300  
|||||  
3241 CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCAGTATTAATGAAA 3300  
|||||  
3301 TAGGTTCCAGTGTAGAAACATTCACAGCAGAACTAGTAGAAACAGAGGCCCCAAATTTGA 3360  
|||||  
3301 TAGGTTCCAGTGTAGAAACATTCACAGCAGAACTAGTAGAAACAGAGGCCCCAAATTTGA 3360  
|||||  
3361 ATGCTATGCTTAGATAGGGGTTTTGCAACCTGAGCTGTATTAACAAAGCTCTCTGTGAA 3420  
|||||  
3361 ATGCTATGCTTAGATAGGGGTTTTGCAACCTGAGGCTGTATTAACAAAGCTCTCTGTGAA 3420  
|||||  
3421 GTAATTTGATGATCCTGTAATTAATAAAGCAAGAAATGAAGATAGTTGACAGCTGTTA 3480  
|||||  
3421 GTAATTTGATGATCCTGTAATTAATAAAGCAAGAAATGAAGATAGTTGACAGCTGTTA 3480  
|||||  
3481 ATACGATTTCTCCCATATCTGATTTAGATTAAGAACAGCTTATGGAAGTATGTC 3540  
|||||  
3481 ATACGATTTCTCCCATATCTGATTTAGATTAAGAACAGCTTATGGAAGTATGTC 3540  
|||||  
3481 ATACGATTTCTCCCATATCTGATTTAGATTAAGAACAGCTTATGGAAGTATGTC 3540  
|||||  
3541 ATGCTATGCTTAGATAGGGGTTTTGCAACCTGAGCTGTATTAACAAAGCTCTCTGTGAA 3600  
|||||  
3541 ATGCTATGCTTAGATAGGGGTTTTGCAACCTGAGGCTGTATTAACAAAGCTCTCTGTGAA 3600  
|||||  
3601 AAGTACTAGTTTGTGCTGAAATGACATTAAGGAAAGTTCTGCTGTTTTAGCAAAAGCG 3660  
|||||  
3601 AAGTACTAGTTTGTGCTGAAATGACATTAAGGAAAGTTCTGCTGTTTTAGCAAAAGCG 3660  
|||||  
3661 TCCGAAAGAGAGCTTAGAGAGAGTCTTACGCCCTTACCCATACACATTTGGCTCAGG 3720  
|||||  
3661 TCCGAAAGAGAGCTTAGAGAGAGTCTTACGCCCTTACCCATACACATTTGGCTCAGG 3720  
|||||  
3721 GTTACCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3780  
|||||  
3721 GTTACCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3780  
|||||  
3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 3840  
|||||  
3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 3840  
|||||  
3841 CTACAGGATAGAGACCTGCTGCTACCGAGTCTGCTCTAAGAACACAGAGAGAGATTTAT 3900  
|||||  
3841 CTACAGGATAGAGACCTGCTGCTACCGAGTCTGCTCTAAGAACACAGAGAGAGATTTAT 3900  
|||||  
3901 TATCATTAAGAAATAGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3960  
|||||  
3901 TATCATTAAGAAATAGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3960  
|||||  
3961 AGGAACATCACCTTAGTGAGGAAACAAATATGTTCTGTAGCTGTTTTCTTACAGTGCA 4020  
|||||

3961 AGGAACATCACCTTAGTGAGGAAACAAATATGTTCTGTAGCTGTTTTCTTACAGTGCA 4020  
|||||  
4021 GTGAATTTGAGAGCTTTGACTGCAATTACAAACCCAGAGATCTTTCTTGATGTTCTT 4080  
|||||  
4021 GTGAATTTGAGAGCTTTGACTGCAATTACAAACCCAGAGATCTTTCTTGATGTTCTT 4080  
|||||  
4081 CCMAACAAATGAGGATCAGTGTGAAGCCAGAGGAGTGGTGTGATGCAAGAAATGG 4140  
|||||  
4081 CCMAACAAATGAGGATCAGTGTGAAGCCAGAGGAGTGGTGTGATGCAAGAAATGG 4140  
|||||  
4141 TTTGATGATGATGAAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4200  
|||||  
4141 TTTGATGATGATGAAGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4200  
|||||  
4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAACAAAGCCCTCTGAA 4260  
|||||  
4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAACAAAGCCCTCTGAA 4260  
|||||  
4261 ACTGCTCAGGGCTATCTCTCTGAGAGTGAACATTTTAAACCTCAGCAGAGGAGTACCATGC 4320  
|||||  
4261 ACTGCTCAGGGCTATCTCTCTGAGAGTGAACATTTTAAACCTCAGCAGAGGAGTACCATGC 4320  
|||||  
4321 AACATTAACCTGATTAAGGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAGAAGCAGC 4380  
|||||  
4321 AACATTAACCTGATTAAGGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAGAAGCAGC 4380  
|||||  
4381 ATGGAGCCAGCTTCTTAACAGCTACCTTCCATATTAAGTATGATGATGATGATGATGATGATG 4440  
|||||  
4381 ATGGAGCCAGCTTCTTAACAGCTACCTTCCATATTAAGTATGATGATGATGATGATGATGATG 4440  
|||||  
4441 ATGGAGCCAGCTTCTTAACAGCTACCTTCCATATTAAGTATGATGATGATGATGATGATGATG 4440  
|||||  
4441 ATGGAGCCAGCTTCTTAACAGCTACCTTCCATATTAAGTATGATGATGATGATGATGATGATG 4440  
|||||  
4441 ACCTCGAAATCCCAACAAAGCAGATCAGAAAGAGAGTATTAACCTCAGCAGAAAGTGA 4500  
|||||  
4441 ACCTCGAAATCCCAACAAAGCAGATCAGAAAGAGAGTATTAACCTCAGCAGAAAGTGA 4500  
|||||  
4441 ACCTCGAAATCCCAACAAAGCAGATCAGAAAGAGAGTATTAACCTCAGCAGAAAGTGA 4500  
|||||  
4501 GTGAATTAACCTGATTAAGGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAGAAGCAGC 4560  
|||||  
4501 GTGAATTAACCTGATTAAGGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAGAAGCAGC 4560  
|||||  
4501 GTGAATTAACCTGATTAAGGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAGAAGCAGC 4560  
|||||  
4561 CAGATAGTTTCAACAGCTTAATAAAGCAAGAGAGTGAAGAGTATGATGATGATGATGATGATG 4620  
|||||  
4561 CAGATAGTTTCAACAGCTTAATAAAGCAAGAGAGTGAAGAGTATGATGATGATGATGATGATG 4620  
|||||  
4561 CAGATAGTTTCAACAGCTTAATAAAGCAAGAGAGTGAAGAGTATGATGATGATGATGATGATG 4620  
|||||  
4621 GCCCATCTTATGATATAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
|||||  
4621 GCCCATCTTATGATATAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
|||||  
4621 GCCCATCTTATGATATAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
|||||  
4681 ACTACCATCTCAAGAGAGGCTCATTAAGGTTGATGATGATGATGATGATGATGATGATGATG 4740  
|||||  
4681 ACTACCATCTCAAGAGAGGCTCATTAAGGTTGATGATGATGATGATGATGATGATGATGATG 4740  
|||||  
4741 AGTCTGGGCGACAGCATTTGAGAGGAAACATCTTACTTCCAAAGGCAAGATCTAGAGGAA 4800  
|||||  
4741 AGTCTGGGCGACAGCATTTGAGAGGAAACATCTTACTTCCAAAGGCAAGATCTAGAGGAA 4800  
|||||  
4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGAGACCTGGAATCTGATCTCTG 4860  
|||||  
4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGAGACCTGGAATCTGATCTCTG 4860  
|||||  
4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGAGACCTGGAATCTGATCTCTG 4860  
|||||  
4861 AAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4920  
|||||  
4861 AAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4920  
|||||  
4861 AAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4920  
|||||  
4921 AAGTTCCCAATTTGAAGTTGAGAGATCTGAGAGTGCAGAGTGCCTCTCATATCTAGT 4980  
|||||  
4921 AAGTTCCCAATTTGAAGTTGAGAGATCTGAGAGTGCAGAGTGCCTCTCATATCTAGT 4980  
|||||  
4981 ATACTGCTGGGTATTAATCAATGAGAGAAAGTGTGAGAGGAGAGGAGGAGGAGGAGGAGGAGG 5040  
|||||  
4981 ATACTGCTGGGTATTAATCAATGAGAGAAAGTGTGAGAGGAGAGGAGGAGGAGGAGGAGGAGG 5040  
|||||  
5041 CTTCAACGAAAGGCTCAACAAAGAAATGTCATGTTGTTGCTGAGCTGAGCCAGAG 5100  
|||||

Db 5041 CTTCAACAGAAAGGCTCAACAAAAGAAATGTCATGATGCTGCTGCGCTGACCCAGAG 5100  
 QY 5101 AATTTATGCTGCTAGCAAGTTTGCAGAAAACACACATCTTAATCTAATCAATTA 5160  
 Db 5101 AATTTATGCTGCTAGCAAGTTTGCAGAAAACACACATCTTAATCTAATCAATTA 5160  
 QY 5161 CTGAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTTGTGTGAGAGGAC 5220  
 Db 5161 CTGAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTTGTGTGAGAGGAC 5220  
 QY 5221 TGAATATTTTCTAGCAATTCGGGAGAGAAAATGGGATGATTTCTGGGTGACC 5280  
 Db 5221 TGAATATTTTCTAGCAATTCGGGAGAGAAAATGGGATGATTTCTGGGTGACC 5280  
 QY 5281 AGCTATTTAAAGAAAAGAAATGCTGATGATGATTTTGAAGTCCAGAGAGATGCG 5340  
 Db 5281 AGCTATTTAAAGAAAAGAAATGCTGATGATGATTTTGAAGTCCAGAGAGATGCG 5340  
 QY 5341 TCAATGGAAGAAACACCAAGGTCAGAGAGAGAGAGAAATCCAGAGAGAAAGATCT 5400  
 Db 5341 TCAATGGAAGAAACACCAAGGTCAGAGAGAGAGAGAAATCCAGAGAGAAAGATCT 5400  
 QY 5401 TCAGGGGGCTAGAAATCTGTTGCTATGGGCCCTTACCACATGCCCAGATCACTGG 5460  
 Db 5401 TCAGGGGGCTAGAAATCTGTTGCTATGGGCCCTTACCACATGCCCAGATCACTGG 5460  
 QY 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5520  
 Db 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5520  
 QY 5521 GCACAGGTGTCACACCAATTTGCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAG 5580  
 Db 5521 GCACAGGTGTCACACCAATTTGCTGTTGTCAGAGAGAGAGAGAGAGAGAGAGAG 5580  
 QY 5581 TCCATGCAATTTGGGCAATGTTGAGGACCTGTTGAGACCCGAGAGTGGTGTGACA 5640  
 Db 5581 TCCATGCAATTTGGGCAATGTTGAGGACCTGTTGAGACCCGAGAGTGGTGTGACA 5640  
 QY 5641 GTGTAGCACTCTACAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5700  
 Db 5641 GTGTAGCACTCTACAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5700  
 QY 5701 GCCACTACTGA 5711  
 Db 5701 GCCACTACTGA 5711  
 RESULT 4  
 HS014680  
 LOCUS  
 DEFINITION Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA,  
 complete cds.  
 ACCESSION U14680  
 VERSION U14680.1 GI:555931  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 5711)  
 Miki,Y., Swensen,J., Shattuck-Eidens,D., Futreal,P.A., Harshman,K.,  
 Tavtigian,S., Liu,Q., Cochran,C., Bennett,L.M., Ding,W., Bell,R.,  
 Rosenthal,J., Hussey,C., Tran,T., McClure,M., Frye,C., Hattier,T.,  
 Phelps,R., Haugen-Strano,A., Katcher,H., Yakumo,K., Gholami,Z.,  
 Shaffer,D., Stone,S., Bayer,S., Wray,C., Bogden,R., Dayananth,P.,  
 Ward,J., Tonin,P., Narod,S., Bristow,P.K., Norris,F.H.,  
 Helvering,L., Morrison,P., Kosteck,P., Lai,M., Barrettl,J.C.,  
 Lewis,C., Neuhausen,S., Cannon-Albright,L., Goldgar,D., Wiseman,R.,  
 Kamb,A. and Skolnick,M.H.  
 A strong candidate for the breast and ovarian cancer susceptibility  
 gene BRCA1  
 JOURNAL Science 266 (5182), 66-71 (1994)  
 MEDLINE 95023896

PUBMED 7545954  
 REFERENCE 2 (bases 1 to 5711)  
 AUTHORS Skolnick,M.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-SEP-1994) Mark H. Skolnick, Myriad Genetics Inc.,  
 the University of Utah, 421 Wakara Way, Suite 201, Salt Lake City,  
 UT 84108, USA  
 FEATURES  
 source  
 Location/Qualifiers  
 1..5711  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q21; spans D17S855"  
 /note="For sequence of alternatively spliced exon 4, see  
 GenBank Accession Number U15595"  
 1..5711  
 /gene="BRCA1"  
 1..119  
 /gene="BRCA1"  
 1..100  
 /gene="BRCA1"  
 /number=1  
 101..199  
 /gene="BRCA1"  
 /number=2  
 120..5711  
 /gene="BRCA1"  
 /note="influences susceptibility to breast and ovarian  
 cancer"  
 /codon\_start=1  
 /product="breast and ovarian cancer susceptibility"  
 /protein\_id="AA03985.1"  
 /db\_xref="GI:555932"  
 /translation="MDLSALRFEVONVNIAMOKILIECPICLIEKEPSTCDHIFC  
 KFCMLKLNKKGPSCPCCKNDITRSLDESTRPSQVLEBELKIIICAFQDITGLIETA  
 NSYNFAKKENNSPEHLKDEYSIIQSMGYRRARRLQSEPEMPSLOETSLVOLNLG  
 TVRLTRKRIQPOKTSVYIELSDSESDVKNKATVCSYGDDELLQITPQGRDEISL  
 DSAKAKACESESDVNTNTEHHQPSNNDLNTTERAARHPKRYOGSSVNLVYPCGT  
 NTHASSLOHENSLLLTQDKRMNKEAFKCSKQPLARSOHNRAGSRETNDRTPT  
 STERKYDNLMDPLCEKRENNKOKLPCSEMDPRDEYPTLINSLOKVNEMRSDEL  
 LGSDDSHDGSSESNKAYADYLVNEDYVSGSEKIDILASDPHALCKSERHSK  
 SVESNIEDKIEFGTYRKASLPNLSHVTENLIIGAVTEPQIIQERPLNKLKRRRP  
 TSGIHPEDFTIKKADLAVOKPPEMNOGTQEDNGQVNNITNSGHENKTKGDSIQNEK  
 NPNPIEISEKSAFTKAEPISSISNMELELHNSKAPKRNLRKSTSHIHLE  
 LVYSRNISPNCTEIOIDSSSEIEIKKKKYNOMPVHRNRLQIMGKRPATGAKKN  
 KPNQOTSRRSDPTPEPKLTNAPGSEFTCSYNSELKEFVNSLSPREEDELEVKY  
 SNNADPKDLMSGERVLOTTERSVSSSILVPGIDYGTQESISLEVSILKRAKTEP  
 NKCVSQCAAFENKGLIHGCSKDNNDTGGFKYPLIGHVNSHRETSIEEESLDAOY  
 LQNTFVKSRQSFAPSPNPGNAEECATFSAHGSLKQSPKVTPECEKEENQCKNE  
 SNIRPVQTVNITAGFPVYQKDKPVDNACSIKSGSRFCLISQFRNGEGLITPKHG  
 LQNPYRIPLPFIKSFVKTKCKKLNLENEFEHSMSPREMGNNIPIVSTISRN  
 IRENVFEASNSNINFGSSTNEVGSINEDISDENIOAEIAGRNGPLNMLPLVY  
 LQPEYVKQSLPGSNCKHPEIKKQYEVEVQVYVTPDSPLISNLEQPMGSSHAQVC  
 SETPDDLLDGELEIKEDTSFANDIKESAFKSVQKGLSRSPFTTHLHAQGR  
 GAKLSESEBENLSDEBELPCFQHLLEGVANNIPSOQSTRSHVATECLSKNEENILS  
 LKNSLNSNOVILIARASQEHLSSETKCSASLPSQCESELDELANTTOPPLIGS  
 SKOMRHOSESQVGLSDKELVSDEREGRGLENNOESOMSDNLGEAGCESEFSV  
 SPQCSGISQSDILITTOQGRDTHNLIKIQOEWAELEAVLEDHGSPMSYSITSDS  
 SALEDLRNPEQSTSEKAVYLISQKSEYRIPSONEGELSAKFEYSDSSSKRKEGVE  
 RSSPSKQPSLDDRWKMSGSLQNRNTRSDREAEKRVNIVVEQDQESBPDLKTES  
 LPQRODLEGTPYLESLSDSESDPSDEPDAEASRVANIPQSLALKPDLAQT  
 AOSPAAHNTDTAGVYNAMESSYREKPELTASTERNKMSMVSQGLPPEEMLYK  
 ARKHNTLTLTLEETHVVMKDAEFVCEERTIKYPLIGAGKMWVSYPVYQSLIKR  
 KMLNHEDFEVRGDVYGRNHOCPKRRARESDOKRIPGLEICCGYGPNTNPQOLEMAY  
 QLCGASVKEHLSFTLGTGYHPLVVVQAPDAWEDNCFHAIQMCAPVYTRKMLVDSV  
 ALQCCDELDTYLLPQIPHSY"  
 200..253  
 /gene="BRCA1"  
 /number=3  
 254..331  
 /gene="BRCA1"  
 /number=5  
 332..420  
 exon  
 exon  
 exon

```

/gene="BRCA1"
/number=6
421..560
/gene="BRCA1"
/number=7
561..665
/gene="BRCA1"
/number=8
666..712
/gene="BRCA1"
/number=9
713..788
/gene="BRCA1"
/number=10
789..4215
/gene="BRCA1"
/number=11
4216..4302
/gene="BRCA1"
/number=12
4303..4476
/gene="BRCA1"
/number=13
4477..4603
/gene="BRCA1"
/number=14
4604..4794
/gene="BRCA1"
/number=15
4795..5105
/gene="BRCA1"
/number=16
5106..5193
/gene="BRCA1"
/number=17
5194..5273
/gene="BRCA1"
/number=18
5274..5310
/gene="BRCA1"
/number=19
5311..5396
/gene="BRCA1"
/number=20
5397..5451
/gene="BRCA1"
/number=21
5452..5526
/gene="BRCA1"
/number=22
5527..5586
/gene="BRCA1"
/number=23
5587..5711
/gene="BRCA1"
/number=24
1956 a 1099 c 1274 g 1382 t
BASE COUNT
ORIGIN
Query Match 99.9%; Score 5707.8; DB 9; Length 5711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 121 TGGATTATCTGCTCTGCGGTGAGAGATACAAATGCTATATGCTATGCAGAAA 180
Qy 181 TCTTAGAGTCCCTGCTGCTGAGTGTATCAAGAACCTGTCTCCCAAGTGTACC 240
Db 181 TCTTAGAGTGTCCATCTGTGTGAGTGTATCAAGAACCTGTCTCCCAAGTGTACC 240
Qy 241 ACATATTTTGCAGAAATTTTGCATGTGAACCTCTCCACAGAGAAAGGCGCTTCACAGT 300
Db 241 ACATATTTTGCAGAAATTTTGCATGTGAACCTCTCCACAGAGAAAGGCGCTTCACAGT 300
Qy 301 GTCCCTTATGTATAGATATATACCAAAAGAGGCTTCAAGAAAGTACGAGATTAGTC 360
Db 301 GTCCCTTATGTATAGATATATACCAAAAGAGGCTTCAAGAAAGTACGAGATTAGTC 360
Qy 361 AACTGTGAGAGCTATTGAAAATCTTGTGCTTTCAGCTTACACAGTTTGAGT 420
Db 361 AACTGTGAGAGCTATTGAAAATCTTGTGCTTTCAGCTTACACAGTTTGAGT 420
Qy 421 ATGCNAACAGCTATATTTTGCAGAAAGAAATATCTCTGTAACATCTAAAGATG 480
Db 421 ATGCNAACAGCTATATTTTGCAGAAAGAAATATCTCTGTAACATCTAAAGATG 480
Qy 481 AAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTGCCAAAGACTTCTACAGATG 540
Db 481 AAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTGCCAAAGACTTCTACAGATG 540
Qy 541 AACCCGAAATCTCTCCCTGAGAGAAACCACTTCAGTCCCACTCTTAACCTTGAA 600
Db 541 AACCCGAAATCTCTCCCTGAGAGAAACCACTTCAGTCCCACTCTTAACCTTGAA 600
Qy 601 CTGTGAGACCTGTGAGACCAAGCAGGATATCAACCTCAAAAGACGCTGTCTCATTTG 660
Db 601 CTGTGAGACCTGTGAGACCAAGCAGGATATCAACCTCAAAAGACGCTGTCTCATTTG 660
Qy 661 AATTGGATCTGATTTCTCTAAGATACCGTTAATAGGCATTTATTCAGTGTGAG 720
Db 661 AATTGGATCTGATTTCTCTAAGATACCGTTAATAGGCATTTATTCAGTGTGAG 720
Qy 721 ATCAAGATTTGTTCAATATCAACCCCTCAAGAGACCGAGATGAATCATGTTGATTCG 780
Db 721 ATCAAGATTTGTTCAATATCAACCCCTCAAGAGACCGAGATGAATCATGTTGATTCG 780
Qy 781 CAAAAAGGCTGCTGTGAATTTTCTGAGAGGATGTATCAAAATCTGATATCATCAAC 840
Db 781 CAAAAAGGCTGCTGTGTGAATTTTCTGAGAGGATGTATCAAAATCTGATATCATCAAC 840
Qy 841 CCAGTAATATGATTTTGAACACCACTGAGAAAGCGTGACGTGAGAGGATCCAGAAAGT 900
Db 841 CCAGTAATATGATTTTGAACACCACTGAGAAAGCGTGACGTGAGAGGATCCAGAAAGT 900
Qy 901 ATCAGGGTATGTTCTTTCAAACTTTGATGTGAGACCACTGTGCAAAATCTCATGCCA 960
Db 901 ATCAGGGTATGTTCTTTCAAACTTTGATGTGAGACCACTGTGCAAAATCTCATGCCA 960
Qy 961 GCTCATTACAGCATGAGAACGACGATTTATATACATCAAGATGATGATGATGAGAAA 1020
Db 961 GCTCATTACAGCATGAGAACGACGATTTATATACATCAAGATGATGATGATGAGAAA 1020
Qy 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCCTGCTTATGCAAGAGAGCCATATACAGAT 1080
Db 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCCTGCTTATGCAAGAGAGCCATATACAGAT 1080
Qy 1081 GGGCTGGAAGTAAAGAAACATGATATAGTAGCGGACATCCAGACAGAAAAAGGATG 1140
Db 1081 GGGCTGGAAGTAAAGAAACATGATATAGTAGCGGACATCCAGACAGAAAAAGGATG 1140
Qy 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGGAATTAACAGAACTGCGATGCT 1200
Db 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGGAATTAACAGAACTGCGATGCT 1200
Qy 1201 CAGAGAATCTTGAAGATAGTGAAGATGTTCTTGGATTAACCTAATATAGCAGATTCAGA 1260
Db 1201 CAGAGAATCTTGAAGATAGTGAAGATGTTCTTGGATTAACCTAATATAGCAGATTCAGA 1260

```



QY	1261	AAAGTTAATGAGTGGTTTTCCAGAGACTGATGAACCTGTAGTTCGATGACTCACATGATG	1320
Db	1261	AAAGTTAATGAGTGGTTTTCCAGAAAGTGAACCTGTAGTTCGATGACTCACATGATG	1320
QY	1321	GGGAGCTCGAATCAAAATGCCAAGTAGAGTGTGAGAGTCTTGAAGAGTGTGAGAGT	1380
Db	1321	GGGAGCTCGAATCAAAATGCCAAGTAGAGTGTGAGAGTCTTGAATGAGGATG	1380
QY	1381	AATATTCGTGTTCTTCAGAGAAAATAGACTTACTGGCCAGTATCCATGAGGCTTTAA	1440
Db	1381	AATATTCGTGTTCTTCAGAGAAAATAGACTTACTGGCCAGTATCCATGAGGCTTTAA	1440
QY	1441	TATGTAAAGTCGAAAGAGTTCACTCCAAATCAGTAGAGATATTTGAGACAAATAT	1500
Db	1441	TATGTAAAGTCGAAAGAGTTCACTCCAAATCAGTAGAGATATTTGAGACAAATAT	1500
QY	1501	TTGGGAAAACCTATGCGAAGGAGAGCGCTCCCACTTAAAGCATGTAAGTCAAAATC	1560
Db	1501	TTGGGAAAACCTATGCGAAGGAGAGCGCTCCCACTTAAAGCATGTAAGTCAAAATC	1560
QY	1561	TATATATAGAGACATTTGTTACTGAGCCACAGATATATCAAGAGCGTCCCTCACAATA	1620
Db	1561	TATATATAGAGACATTTGTTACTGAGCCACAGATATATCAAGAGCGTCCCTCACAATA	1620
QY	1621	AATTTAAAGCTTAAAGAGACCTACATCAGGCTTCATCTGAGAGTTTATCAGAAG	1680
Db	1621	AATTTAAAGCTTAAAGAGACCTACATCAGGCTTCATCTGAGAGTTTATCAGAAG	1680
QY	1681	CAGATTTGGCAGTTCCAAAAGACTCCTGAAATGATTAATCAGGAACTTACCAAGGAGC	1740
Db	1681	CAGATTTGGCAGTTCCAAAAGACTCCTGAAATGATTAATCAGGAACTTACCAAGGAGC	1740
QY	1741	AGAAATGTCAGTGTGATGATATTTACTTAATAGTGTGTCATGAGATTAACCAAAAGGTAT	1800
Db	1741	AGAAATGTCAGTGTGATGATATTTACTTAATAGTGTGTCATGAGATTAACCAAAAGGTAT	1800
QY	1801	CTATTCAGATGAGAAAAATCTCTAACCCAAATAGAACTCTGGAAGAAATCTGCTTTCA	1860
Db	1801	CTATTCAGATGAGAAAAATCTCTAACCCAAATAGAACTCTGGAAGAAATCTGCTTTCA	1860
QY	1861	AAACGAAAGCTGAACTATTAAGCAGCAGTATTAAGCAATATGGAATCGAATTAATATCC	1920
Db	1861	AAACGAAAGCTGAACTATTAAGCAGCAGTATTAAGCAATATGGAATCGAATTAATATCC	1920
QY	1921	ACAATTCAAAAGCACCCTAAAAAGATAGGCTGAGAGAGAACTCTTCACGAGCATATTC	1980
Db	1921	ACAATTCAAAAGCACCCTAAAAAGATAGGCTGAGAGAGAACTCTTCACGAGCATATTC	1980
QY	1981	ATGGCCTTGAACTATGACTAGTGAAGTAATTTAAAGCCCAATTTGTACTGAATTTGCAAA	2040
Db	1981	ATGGCCTTGAACTATGACTAGTGAAGTAATTTAAAGCCCAATTTGTACTGAATTTGCAAA	2040
QY	2041	TTGATAGTGTGTTCTGACAGTAGAGATTAAGAAAAAAATTTACACCAATTTGCAGTCA	2100
Db	2041	TTGATAGTGTGTTCTGACAGTAGAGATTAAGAAAAAAATTTACACCAATTTGCAGTCA	2100
QY	2101	GGCAGCAGCAAACTTACAACTCATGTGAGAGTAAAGAACTGCAACTGAGACCAAGAGA	2160
Db	2101	GGCAGCAGCAAACTTACAACTCATGTGAGAGTAAAGAACTGCAACTGAGACCAAGAGA	2160
QY	2161	GTAACCAAGCCAAATGACAGACAAAGTAATAACACATGACAGTGAATCTTCCCAAGCTGA	2220
Db	2161	GTAACCAAGCCAAATGACAGACAAAGTAATAACACATGACAGTGAATCTTCCCAAGCTGA	2220
QY	2221	AGTTAACCAATGACAGCTGTTCTTTTACTAGTGTCTAAATATCCAGTGAATTTAAAGAT	2280
Db	2221	AGTTAACCAATGACAGCTGTTCTTTTACTAGTGTCTAAATATCCAGTGAATTTAAAGAT	2280
QY	2281	TTGTCAATCTTACGCTTCCAAAGAGAGAAAGAGAAACTTGAAGACGTTAAAGGT	2340
Db	2281	TTGTCAATCTTACGCTTCCAAAGAGAGAAAGAGAAACTTGAAGACGTTAAAGGT	2340

QY	2341	CTAATAATGCTGAACACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAACTG	2400
Db	2341	CTAAATAATGCTGAACACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAACTG	2400
QY	2401	AAAGATCTGTAGACAGTAGCAGTATTGTGATGGTACCTGACTATTTATGGAACCTCAG	2460
Db	2401	AAAGATCTGTAGACAGTAGCAGTATTGTGATGGTACCTGACTATTTATGGAACCTCAG	2460
QY	2461	AAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAGGCAAAAACAGACCAATAAT	2520
Db	2461	AAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAGGCAAAAACAGACCAATAAT	2520
QY	2521	GTGTAGTCAGTGTGCAGCATTTGAAACCCCAAGGACCTAATTCATGGTTGTTCCAAAG	2580
Db	2521	GTGTAGTCAGTGTGCAGCATTTGAAACCCCAAGGACCTAATTCATGGTTGTTCCAAAG	2580
QY	2581	ATATAAGAAATGCACACAGAGGCTTTAAGTATCCATTGGGACATGAAATTAACCACTG	2640
Db	2581	ATATAAGAAATGCACACAGAGGCTTTAAGTATCCATTGGGACATGAAATTAACCACTG	2640
QY	2641	GGGAAACACAGCATAGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGCAGATACAT	2700
Db	2641	GGGAAACACAGCATAGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGCAGATACAT	2700
QY	2701	TCAAGGTTTCAAAAGGCGCAGTCATTGGCTGTGTTTCAAAATCCAGAAATCAGAAAGG	2760
Db	2701	TCAAGGTTTCAAAAGGCGCAGTCATTGGCTGTGTTTCAAAATCCAGAAATCAGAAAGG	2760
QY	2761	AATGTGCACATCTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAGCTACT	2820
Db	2761	AATGTGCACATCTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAGCTACT	2820
QY	2821	TTGATGTGACACAAAGGAAGAAATCAAGGAAGAATGAGCTAATTCAGGCTGTAC	2880
Db	2821	TTGATGTGACACAAAGGAAGAAATCAAGGAAGAATGAGCTAATTCAGGCTGTAC	2880
QY	2881	AGACAGTTAATATCATCTGACAGGCTTCCGTGGTGGTGCAGAAAGATAAAGCAGTTGATA	2940
Db	2881	AGACAGTTAATATCATCTGACAGGCTTCCGTGGTGGTGCAGAAAGATAAAGCAGTTGATA	2940
QY	2941	ATGCCAATGTAGTTCAAAGAGGCTCTAGGTTTGTCTTCATCTCAGTTCAGAGGCA	3000
Db	2941	ATGCCAATGTAGTTCAAAGAGGCTCTAGGTTTGTCTTCATCTCAGTTCAGAGGCA	3000
QY	3001	ACGAATACGAGTCACTACTCCAAATAAACATGAGACTTTTCAAAACCATATGCTATAC	3060
Db	3001	ACGAATACGAGTCACTACTCCAAATAAACATGAGACTTTTCAAAACCATATGCTATAC	3060
QY	3061	CACCACTTTTCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCTGCTAGAGC	3120
Db	3061	CACCACTTTTCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCTGCTAGAGC	3120
QY	3121	AAAACCTTGAGAGACATTCATGTCACCTGAAAGAAATGGAAATGAGAAACATTCCAA	3180
Db	3121	AAAACCTTGAGAGACATTCATGTCACCTGAAAGAAATGGAAATGAGAAACATTCCAA	3180
QY	3181	GTACAGTAGACCAATTAAGCGTAAATACATTAGAGAAAAATGTTTTAAAGAACCCAGCT	3240
Db	3181	GTACAGTAGACCAATTAAGCGTAAATACATTAGAGAAAAATGTTTTAAAGAACCCAGCT	3240
QY	3241	CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGGTGGGCTCCAGATTAATGAAA	3300
Db	3241	CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGGTGGGCTCCAGATTAATGAAA	3300
QY	3301	TAGGTTCCAGTAGAAGAAACATTCAGAGAACTAGTAGAAACAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTAGAAGAAACATTCAGAGAACTAGTAGAAACAGAGGGCCAAATTTGA	3360
QY	3361	ATGCTATGCTTAGATTAGGGGTTTTGCACCTGAGGCTCTAATAAACAAAGTCTTCTCGGAA	3420
Db	3361	ATGCTATGCTTAGATTAGGGGTTTTGCACCTGAGGCTCTAATAAACAAAGTCTTCTCGGAA	3420
QY	3421	GTAATTGTAAACATCTGTAATAAATAAAGCAAGATTTGAAGAGTGTGTTCAAGCTGTA	3480



|||||  
 Db 3421 GTAATGTAGATCCTGAAATTAAGCAAGATATGAGAGTACTGACGTGTA 3480  
 Oy 3481 ATACGATTTCTCTCCATATCTGATTTAGATTAAGTGAACGCTTATGGAGTACTG 3540  
 Db 3481 ATACGATTTCTCTCCATATCTGATTTAGATTAAGTGAACGCTTATGGAGTACTG 3540  
 Oy 3541 ATGATCTCAGGTTGTCTGTGAGACCTGATGACCTTTAGATGATGATGATGATGATG 3600  
 Db 3541 ATGATCTCAGGTTGTCTGTGAGACCTGATGACCTTTAGATGATGATGATGATGATG 3600  
 Oy 3601 AAGATCTAGTTTGTCTGATTAAGTGAAGTGTCTGTTTATGCAAAAGCG 3660  
 Db 3601 AAGATCTAGTTTGTCTGATTAAGTGAAGTGTCTGTTTATGCAAAAGCG 3660  
 Oy 3661 TCCAGAAAGAGAGCTTGAAGAGAGTCTAGACCTTTACCATTAACATTTGGCTCAGG 3720  
 Db 3661 TCCAGAAAGAGAGCTTGAAGAGAGTCTAGACCTTTACCATTAACATTTGGCTCAGG 3720  
 Oy 3721 GTTACCGAAGAGGGGCCAAGAAATAGAGTCTCAGAGAGAACTTATCTAGTGAAGT 3780  
 Db 3721 GTTACCGAAGAGGGGCCAAGAAATAGAGTCTCAGAGAGAACTTATCTAGTGAAGT 3780  
 Oy 3781 AAGAGCTTCCCTGCTTCCACACTGTTATTTGTTAAAGTAAACAATATACCTTCTCAGT 3840  
 Db 3781 AAGAGCTTCCCTGCTTCCACACTGTTATTTGTTAAAGTAAACAATATACCTTCTCAGT 3840  
 Oy 3841 CTACAGGATAGACACCTGCTGCTACCGAGTCTCTCTAAGAACACAGAGAGATTTAT 3900  
 Db 3841 CTACAGGATAGACACCTGCTGCTACCGAGTCTCTCTAAGAACACAGAGAGATTTAT 3900  
 Oy 3901 TATCATTAAGAAATAGTTAATGATGATGATGATGATGATGATGATGATGATGATG 3960  
 Db 3901 TATCATTAAGAAATAGTTAATGATGATGATGATGATGATGATGATGATGATGATG 3960  
 Oy 3961 AGGAACATCACCTTATGATGAGAAACAAATGTTCTGATGATGATGATGATGATGATG 4020  
 Db 3961 AGGAACATCACCTTATGATGAGAAACAAATGTTCTGATGATGATGATGATGATGATG 4020  
 Oy 4021 GTGATTTGGAAGACTGCTGCTCAAAATACAAACACCCAGAGATCTTTCTGATGATG 4080  
 Db 4021 GTGATTTGGAAGACTGCTGCTCAAAATACAAACACCCAGAGATCTTTCTGATGATG 4080  
 Oy 4081 CCAAAACAATAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 4140  
 Db 4081 CCAAAACAATAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 4140  
 Oy 4141 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4200  
 Db 4141 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4200  
 Oy 4201 TGGATTTCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 4260  
 Db 4201 TGGATTTCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 4260  
 Oy 4261 ACTGCTCAGGCTATCTCTCAGAGTGAATTTTAACTCAGAGAGAGGATACCATG 4320  
 Db 4261 ACTGCTCAGGCTATCTCTCAGAGTGAATTTTAACTCAGAGAGAGGATACCATG 4320  
 Oy 4321 AACATTAACCTGATTAAGCTTCCAGAGAAATGGCTTGAAGTGAAGTGTGTTAGAACG 4380  
 Db 4321 AACATTAACCTGATTAAGCTTCCAGAGAAATGGCTTGAAGTGAAGTGTGTTAGAACG 4380  
 Oy 4381 ATGGAGAGGAGGCTTCTTAAGAGTACCTTCCATCATTAAGTGAAGTGTGTTAGAACG 4440  
 Db 4381 ATGGAGAGGAGGCTTCTTAAGAGTACCTTCCATCATTAAGTGAAGTGTGTTAGAACG 4440  
 Oy 4441 ACCTGCGAAATCCAGAACAAAGACATCAGAAAAAGAGTATTAACCTCAGAGAAAGTA 4500  
 Db 4441 ACCTGCGAAATCCAGAACAAAGACATCAGAAAAAGAGTATTAACCTCAGAGAAAGTA 4500  
 Oy 4501 GTGAATACCTTAAGAGGAGAGTCCAGAAAGGCTTTCTGATGATGATGATGATGATG 4560  
 Db 4501 GTGAATACCTTAAGAGGAGAGTCCAGAAAGGCTTTCTGATGATGATGATGATGATG 4560

Db 4501 GTGAATACCTTAAGAGGAGAGTCCAGAAAGGCTTTCTGATGATGATGATGATGATG 4560  
 Oy 4561 CAGATAGTTCTACCGATTAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
 Db 4561 CAGATAGTTCTACCGATTAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
 Oy 4621 GCCCATCATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
 Db 4621 GCCCATCATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
 Oy 4681 ACTACCATCTCCAGAGAGAGTCAATTAAGTGTGATGATGATGATGATGATGATGATG 4740  
 Db 4681 ACTACCATCTCCAGAGAGAGTCAATTAAGTGTGATGATGATGATGATGATGATGATG 4740  
 Oy 4741 AGTCTGGGCGACAGCATTTGAGAGAAACATCTTACTTGGCAAGGCAAGATCTAGAGG 4800  
 Db 4741 AGTCTGGGCGACAGCATTTGAGAGAAACATCTTACTTGGCAAGGCAAGATCTAGAGG 4800  
 Oy 4801 CCCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGATGATGATGATGATGATGATG 4860  
 Db 4801 CCCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGATGATGATGATGATGATGATG 4860  
 Oy 4861 AAGACAGAGCCCGAGAGTCAAGTCTGTTGSCAACAATACATCTTCAACCTCTGATG 4920  
 Db 4861 AAGACAGAGCCCGAGAGTCAAGTCTGTTGSCAACAATACATCTTCAACCTCTGATG 4920  
 Oy 4921 AAGTCCCCCAATTTGAAGTGTGAGAAATCTGCTGCAAGTCAAGTCAAGTCAAGT 4980  
 Db 4921 AAGTCCCCCAATTTGAAGTGTGAGAAATCTGCTGCAAGTCAAGTCAAGTCAAGT 4980  
 Oy 4981 ATACTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
 Db 4981 ATACTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
 Oy 4981 ATACTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
 Db 4981 ATACTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
 Oy 5041 CTTCACAGAAAGGCTCAACAAAGATCTCCATGATGATGATGATGATGATGATGATG 5100  
 Db 5041 CTTCACAGAAAGGCTCAACAAAGATCTCCATGATGATGATGATGATGATGATGATG 5100  
 Oy 5101 AATTTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5160  
 Db 5101 AATTTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5160  
 Oy 5161 CTGAGAGACTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5220  
 Db 5161 CTGAGAGACTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5220  
 Oy 5221 TGAATATTTTCTAGGAATTTGGGAGAGAAATGAGTATGATGATGATGATGATGATG 5280  
 Db 5221 TGAATATTTTCTAGGAATTTGGGAGAGAAATGAGTATGATGATGATGATGATGATG 5280  
 Oy 5281 AGTCTATTAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 5340  
 Db 5281 AGTCTATTAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 5340  
 Oy 5341 TCAATGGAAGAAACCAAGAGTCCAAAGCGAGAGAGAAATCCAGAGAGAGAGATG 5400  
 Db 5341 TCAATGGAAGAAACCAAGAGTCCAAAGCGAGAGAGAAATCCAGAGAGAGAGATG 5400  
 Oy 5401 TCAGGGGCTAGAAATCTGTTGCTATGAGGCTTACCAACATGCTCCAGATCAATG 5460  
 Db 5401 TCAGGGGCTAGAAATCTGTTGCTATGAGGCTTACCAACATGCTCCAGATCAATG 5460  
 Oy 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5520  
 Db 5461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5520  
 Oy 5521 GCACAGGCTGACCAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 5580  
 Db 5521 GCACAGGCTGACCAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 5580  
 Oy 5581 TCCATGCAATTTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 5640  
 Db 5581 TCCATGCAATTTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 5640

QY 5641 GTGTACACCTCTACACAGTGGCAGAGCTGACACCTACTGATATCCAGATCCCCACA 5700  
|||||  
Db 5641 GTGTACACCTCTACACAGTGGCAGAGCTGACACCTACTGATATCCAGATCCCCACA 5700  
QY 5701 GCCACTACTGA 5711  
|||||  
Db 5701 GCCACTACTGA 5711  
RESULT 5  
AR070223 5712 bp DNA linear PAT 18-FEB-2000  
LOCUS AR070223  
DEFINITION Sequence 1 from patent US 5891857.  
ACCESSION AR070223  
VERSION AR070223.1 GI:7221111  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 5712)  
AUTHORS Holt,J.T., Jensen,R.A., King,M.-C., Page,D.L., Szabo,C.I.,  
Jettion,T.L., Robinson-Benion,C.L. and Thompson,M.E.  
TITLE Characterized BRCA1 and BRCA2 proteins and screening and  
therapeutic methods based on characterized BRCA1 and BRCA2 proteins  
JOURNAL Patent: US 5891857-A 1 06-Apr-1999;  
FEATURES Location/Qualifiers  
source 1. 5712  
BASE COUNT 1956 a 1099 c 1274 g 1383 t  
ORIGIN  
Query Match 99.9%; Score 5707.8; DB 6; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

|||||  
Db 541 AACCCGAAATCCCTCTCGAGGAACAGCTCAGTGTCCAACTCTTAACCTGGAA 600  
CTGTGACAACTCTGAGGACAAAGAGGGGATACAACTCAAAAGAGCTGTCTACATTG 660  
|||||  
Db 601 CTGTGACAACTCTGAGGACAAAGAGGGGATACAACTCAAAAGAGCTGTCTACATTG 660  
601 CTGTGACAACTCTGAGGACAAAGAGGGGATACAACTCAAAAGAGCTGTCTACATTG 660  
QY 661 AATTGGATCTGATCTTCTGAAAGATACCTTAATAAGCACTTATTCAGAGTGGAG 720  
|||||  
Db 661 AATTGGATCTGATCTTCTGAAAGATACCTTAATAAGCACTTATTCAGAGTGGAG 720  
QY 721 ATCAAGAAATTGTACAATAACCCCTCAGAGAACCAAGGATGAATCACTTGGATTCTG 780  
|||||  
Db 721 ATCAAGAAATTGTACAATAACCCCTCAGAGAACCAAGGATGAATCACTTGGATTCTG 780  
721 ATCAAGAAATTGTACAATAACCCCTCAGAGAACCAAGGATGAATCACTTGGATTCTG 780  
QY 781 CAAAAAGGCTGCTTGATTTTCTGAGAGCGGATGTAACAAATCTGAACATCATCAAC 840  
|||||  
Db 781 CAAAAAGGCTGCTTGATTTTCTGAGAGCGGATGTAACAAATCTGAACATCATCAAC 840  
841 CCAATTAATTAATGATTTGAACACCACTGAGAGCGTCACTGAGAGCATCCAGAAAGT 900  
|||||  
Db 841 CCAATTAATTAATGATTTGAACACCACTGAGAGCGTCACTGAGAGCATCCAGAAAGT 900  
QY 901 ATCAGGCTAGTCTGTTTCAAACTTGATGTGAGCGCATGTGGCACAATTAATCATGCA 960  
|||||  
Db 901 ATCAGGCTAGTCTGTTTCAAACTTGATGTGAGCGCATGTGGCACAATTAATCATGCA 960  
901 ATCAGGCTAGTCTGTTTCAAACTTGATGTGAGCGCATGTGGCACAATTAATCATGCA 960  
QY 961 GCTCATTAACAGCATGAGAACAGAGATTTTACTCATTAAGACAAATGAATGTAGAA 1020  
|||||  
Db 961 GCTCATTAACAGCATGAGAACAGAGATTTTACTCATTAAGACAAATGAATGTAGAA 1020  
961 GCTCATTAACAGCATGAGAACAGAGATTTTACTCATTAAGACAAATGAATGTAGAA 1020  
QY 1021 AGGCTGAATTCCTGTAATAAAGCAACAGCTGCTTACAGAGGACCAATTAACAGAT 1080  
|||||  
Db 1021 AGGCTGAATTCCTGTAATAAAGCAACAGCTGCTTACAGAGGACCAATTAACAGAT 1080  
1021 AGGCTGAATTCCTGTAATAAAGCAACAGCTGCTTACAGAGGACCAATTAACAGAT 1080  
QY 1081 GGGCTGGAAGTANGAAGAACTGTAATGATAGCGGACTCCACAGACAGAAAAAAGTAG 1140  
|||||  
Db 1081 GGGCTGGAAGTANGAAGAACTGTAATGATAGCGGACTCCACAGACAGAAAAAAGTAG 1140  
1081 GGGCTGGAAGTANGAAGAACTGTAATGATAGCGGACTCCACAGACAGAAAAAAGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAGCAAGCAATGCTT 1200  
|||||  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAGCAAGCAATGCTT 1200  
1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAGCAAGCAATGCTT 1200  
QY 1201 CAGAGAACTCTAGAGATGATGAAGATGTCCTTGATTAACACTTAATTAACAGATTGCA 1260  
|||||  
Db 1201 CAGAGAACTCTAGAGATGATGAAGATGTCCTTGATTAACACTTAATTAACAGATTGCA 1260  
1201 CAGAGAACTCTAGAGATGATGAAGATGTCCTTGATTAACACTTAATTAACAGATTGCA 1260  
QY 1261 AAGTTAATGAGTGTGTTTCCAGAGATGATGACTGTTAGCTTCTGATGACTCATGATG 1320  
|||||  
Db 1261 AAGTTAATGAGTGTGTTTCCAGAGATGATGACTGTTAGCTTCTGATGACTCATGATG 1320  
1261 AAGTTAATGAGTGTGTTTCCAGAGATGATGACTGTTAGCTTCTGATGACTCATGATG 1320  
QY 1321 GGGAGTCTGAATCAAAATGCCAAAGTACGATGATTTGGACGCTCAAAATGAGTAGATG 1380  
|||||  
Db 1321 GGGAGTCTGAATCAAAATGCCAAAGTACGATGATTTGGACGCTCAAAATGAGTAGATG 1380  
1321 GGGAGTCTGAATCAAAATGCCAAAGTACGATGATTTGGACGCTCAAAATGAGTAGATG 1380  
QY 1381 AATATTCTGGTCTTCAGAGAAATAGACTTACTGGCCAGTATCCCTATGAGGCTTTAA 1440  
|||||  
Db 1381 AATATTCTGGTCTTCAGAGAAATAGACTTACTGGCCAGTATCCCTATGAGGCTTTAA 1440  
1381 AATATTCTGGTCTTCAGAGAAATAGACTTACTGGCCAGTATCCCTATGAGGCTTTAA 1440  
QY 1441 TATGTAAAAGTGAAGAGTTCACTCCAATTCAGTAGAGATTAATTTGAGACAAATAT 1500  
|||||  
Db 1441 TATGTAAAAGTGAAGAGTTCACTCCAATTCAGTAGAGATTAATTTGAGACAAATAT 1500  
1441 TATGTAAAAGTGAAGAGTTCACTCCAATTCAGTAGAGATTAATTTGAGACAAATAT 1500  
QY 1501 TTGGGAAAACCTATGGGAAGAGGCAAGGCTCCCAACTTAAGCCATGATCAATGAANAAT 1560  
|||||  
Db 1501 TTGGGAAAACCTATGGGAAGAGGCAAGGCTCCCAACTTAAGCCATGATCAATGAANAAT 1560  
1501 TTGGGAAAACCTATGGGAAGAGGCAAGGCTCCCAACTTAAGCCATGATCAATGAANAAT 1560  
QY 1561 TAATTTAGAGAGATTTGTTACTGAGCCACAGATATAACAAAGAGGTCCTCCACAAATA 1620  
|||||  
Db 1561 TAATTTAGAGAGATTTGTTACTGAGCCACAGATATAACAAAGAGGTCCTCCACAAATA 1620  
1561 TAATTTAGAGAGATTTGTTACTGAGCCACAGATATAACAAAGAGGTCCTCCACAAATA 1620  
QY 1621 AATTAAGCGTAAAGAGACCTACATCAGGCTTCATCTGAGGATTTTATCAAGAAAG 1680  
|||||

Dd 1621 AATTAAAGCTTAAAGAGACCTACATCAGGCTTCATCTGAGATTTCACAAAG 1680  
Qy 1681 CAGATTGGCAGTTCAAAGACTCCTGAAATGATTAATCAGGAACTAACCAAGGAGC 1740  
Dd 1681 CAGATTGGCAGTTCAAAGACTCCTGAAATGATTAATCAGGAACTAACCAAGGAGC 1740  
Qy 1741 AGAATGGTCAAGTGAATATTTACTATAGTGGTCATGAGAAATAAACAAAGTGAT 1800  
Dd 1741 AGAATGGTCAAGTGAATATTTACTATAGTGGTCATGAGAAATAAACAAAGTGAT 1800  
Qy 1801 CTATTCAGATGAGAAAAATCTTAACCAATAGATCTCTGAAAAAGATCTGTTTCA 1860  
Dd 1801 CTATTCAGAAATGAGAAAAATCTTAACCAATAGATCTCTGAAAAAGATCTGTTTCA 1860  
Qy 1861 AAACGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAATGCAATTAATATCC 1920  
Dd 1861 AAACGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAATGCAATTAATATCC 1920  
Qy 1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGGAGAGAGCTTCTACAGGATATTC 1980  
Dd 1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGGAGAGAGCTTCTACAGGATATTC 1980  
Qy 1981 ATGCGCTTGAACCTAGTACTGATGAAATCTAACCCACCTAATTTGTAATTCGAAA 2040  
Dd 1981 ATGCGCTTGAACCTAGTACTGATGAAATCTAACCCACCTAATTTGTAATTCGAAA 2040  
Qy 2041 TTGATAGTGTTCAGAGTGAAGATTAAGAAAAAAGTACACCAATGGCAGCA 2100  
Dd 2041 TTGATAGTGTTCAGAGTGAAGATTAAGAAAAAAGTACACCAATGGCAGCA 2100  
Qy 2101 GGCACAGCAGAAACCTCACTCACTGAGAGTAAAGAACTGCACTGAGCAGCAAGA 2160  
Dd 2101 GGCACAGCAGAAACCTCACTCACTGAGAGTAAAGAACTGCACTGAGCAGCAAGA 2160  
Qy 2161 GTAACAAACCCAAATGAGACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
Dd 2161 GTAACAAACCCAAATGAGACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
Qy 2221 AGTTAAACAAATGACAGCTGTTCTTTACTAGTGTCAATATGAGTAAAGAT 2280  
Dd 2221 AGTTAAACAAATGACAGCTGTTCTTTACTAGTGTCAATATGAGTAAAGAT 2280  
Qy 2281 TTGTCAATCTCCTAGCTTCCAGAGAGAAAGAAAGAACTAGAAACACTTAAAGTGT 2340  
Dd 2281 TTGTCAATCTCCTAGCTTCCAGAGAGAAAGAAAGAACTAGAAACACTTAAAGTGT 2340  
Qy 2341 CTATTAATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAACTG 2400  
Dd 2341 CTATTAATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAACTG 2400  
Qy 2401 AAAGATCTGTAGAGATGAGCATTTTCATGTCACCTGAGTATTAAGGCACTCAGG 2460  
Dd 2401 AAAGATCTGTAGAGATGAGCATTTTCATGTCACCTGAGTATTAAGGCACTCAGG 2460  
Qy 2461 AAAGATCTGTAGAGATGAGCATTTTCATGTCACCTGAGTATTAAGGCACTCAGG 2520  
Dd 2461 AAAGATCTGTAGAGATGAGCATTTTCATGTCACCTGAGTATTAAGGCACTCAGG 2520  
Qy 2521 GTGTGAGTCACTGTGTCAGCATTTTAAACCCCAAGGAGCTAATTCATGTTTTCGAAAG 2580  
Dd 2521 GTGTGAGTCACTGTGTCAGCATTTTAAACCCCAAGGAGCTAATTCATGTTTTCGAAAG 2580  
Qy 2581 ATATAGAAATGAGACAGAGGCTTTAGTATTCATGTCAGGAGCTAAGGTTTACCAAGTC 2640  
Dd 2581 ATATAGAAATGAGACAGAGGCTTTAGTATTCATGTCAGGAGCTAAGGTTTACCAAGTC 2640  
Qy 2641 GGGAAACAGAGATGAATGAGAAAGTGAATTCAGTATTCAGTATTCAGTATTCAGT 2700  
Dd 2641 GGGAAACAGAGATGAATGAGAAAGTGAATTCAGTATTCAGTATTCAGTATTCAGT 2700  
Qy 2701 TCAAGGTTTCAAGGCGCAGTATTCGTTCTGTTTCAATTCAGGAAATGCAAGAGG 2760  
Dd 2701 TCAAGGTTTCAAGGCGCAGTATTCGTTCTGTTTCAATTCAGGAAATGCAAGAGG 2760

Qy 2761 AATGTGCAACATTCCTCTGCCCCACCTGCGCTTTAAAGAAACAAAGTCCAAAGTCACT 2820  
Dd 2761 AATGTGCAACATTCCTCTGCCCCACCTGCGCTTTAAAGAAACAAAGTCCAAAGTCACT 2820  
Qy 2821 TTGAATGTGAACAAAGAGAGAAATTCAGAGAAAGATGATTAATATCAAGCTGTAC 2880  
Dd 2821 TTGAATGTGAACAAAGAGAGAAATTCAGAGAAAGATGATTAATATCAAGCTGTAC 2880  
Qy 2881 AGACAGTTAATATCAGTCTGAGGCTTCTGTTGGTTGGTTCAGAAAGATTAAGCAGTTGATA 2940  
Dd 2881 AGACAGTTAATATCAGTCTGAGGCTTCTGTTGGTTGGTTCAGAAAGATTAAGCAGTTGATA 2940  
Qy 2941 ATGCCAAATGTAGTATCAAGAGAGCTGAGTGTTCCTATTCATCTGAGTCAAGAGCA 3000  
Dd 2941 ATGCCAAATGTAGTATCAAGAGAGCTGAGTGTTCCTATTCATCTGAGTCAAGAGCA 3000  
Qy 3001 ACGAAACTGACATATTCCTCAAAATTAACATGACACTTTTACAAACCCATATCTATAC 3060  
Dd 3001 ACGAAACTGACATATTCCTCAAAATTAACATGACACTTTTACAAACCCATATCTATAC 3060  
Qy 3061 CACCACCTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3120  
Dd 3061 CACCACCTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3120  
Qy 3121 AAACTTTGAGAGACATTCATGTCACCTGAAAGAAAGTGAAGAAATGAGAAACATTCGAA 3180  
Dd 3121 AAACTTTGAGAGACATTCATGTCACCTGAAAGAAAGTGAAGAAATGAGAAACATTCGAA 3180  
Qy 3181 GTACAGTGAAGCAATTAAGCGGTATTAATAGAGAAATGTTTAAAGAGCAGCT 3240  
Dd 3181 GTACAGTGAAGCAATTAAGCGGTATTAATAGAGAAATGTTTAAAGAGCAGCT 3240  
Qy 3241 CAAGCAATATTAATGAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAAA 3300  
Dd 3241 CAAGCAATATTAATGAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAAA 3300  
Qy 3301 TAGGTTCCAGTGAAGAAACATTCAGCAGAACTAGTGAAGAAACAGAGGCGCAAAATGGA 3360  
Dd 3301 TAGGTTCCAGTGAAGAAACATTCAGCAGAACTAGTGAAGAAACAGAGGCGCAAAATGGA 3360  
Qy 3361 ATGCTATGCTTAGATTAAGGCTTTTTCAGCTGAGGCTATTAACAAAGCTCTCTCGAA 3420  
Dd 3361 ATGCTATGCTTAGATTAAGGCTTTTTCAGCTGAGGCTATTAACAAAGCTCTCTCGAA 3420  
Qy 3421 GTAATGTGACATCTCTAAATTAAGAAAGCAAGATATGAAGAACTAGTACAGCTGTA 3480  
Dd 3421 GTAATGTGACATCTCTAAATTAAGAAAGCAAGATATGAAGAACTAGTACAGCTGTA 3480  
Qy 3481 ATACAGATTTCTGTCATATCTGATTTAGATTAAGTGAACAGGCTATGGAAGTATGC 3540  
Dd 3481 ATACAGATTTCTGTCATATCTGATTTAGATTAAGTGAACAGGCTATGGAAGTATGC 3540  
Qy 3541 ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAGG 3600  
Dd 3541 ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAGG 3600  
Qy 3601 AAGATTAAGTGTGCTGAGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAGCG 3660  
Dd 3601 AAGATTAAGTGTGCTGAGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAGCG 3660  
Qy 3661 TCCAGAAAGGAGACCTTAGCAGAGGCTCTACCCCTTACCCATACACATTTGGCTCAGG 3720  
Dd 3661 TCCAGAAAGGAGACCTTAGCAGAGGCTCTACCCCTTACCCATACACATTTGGCTCAGG 3720  
Qy 3721 GTTACGGAAGAGGAGGAGAAATTAAGATCTCAGAGAGAACTTATCTAGAGAGTGC 3780  
Dd 3721 GTTACGGAAGAGGAGGAGAAATTAAGATCTCAGAGAGAACTTATCTAGAGAGTGC 3780  
Qy 3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTAAGATTAACATTAATCTCTCAGT 3840  
Dd 3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTAAGATTAACATTAATCTCTCAGT 3840



source 1. 5712 /organism="unknown"  
BASE COUNT 1956 a 1099 c 1274 g 1383 t  
ORIGIN  
Query Match 99.98; Score 5707.8; DB 6; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 AGCTGCGTGAAGCTTCCGAGCCCGGACCGAGGCTGTGGGTTCTCAGTAACCTGGGCC 60  
DB 1 AGCTGCGTGAAGCTTCCGAGCCCGGACCGAGGCTGTGGGTTCTCAGTAACCTGGGCC 60  
OY 61 CCGTGGCTCAGAGAGGCTTCAACCTGTGCTGGGTAAGTTCAATTGGAGAGAGAAA 120  
DB 61 CCGTGGCTCAGAGAGGCTTCAACCTGTGCTGGGTAAGTTCAATTGGAGAGAGAAA 120  
OY 121 TGAATTTATCTGCTCTTCCGCTTGAAGAAGTACAAATGTCTATTATGCTATGAGAAA 180  
DB 121 TGAATTTATCTGCTCTTCCGCTTGAAGAAGTACAAATGTCTATTATGCTATGAGAAA 180  
OY 181 TCTTAGAGTCCCATGCTGCTGAGTTGATCAAGGAACCTGTCCCAAAAGTGTACC 240  
DB 181 TCTTAGAGTCCCATGCTGCTGAGTTGATCAAGGAACCTGTCCCAAAAGTGTACC 240  
OY 241 ACATATTTTGGCAATTTTGCATGCTGAACCTTCTCAACAGAGAGAGAGGCTTCAACAGT 300  
DB 241 ACATATTTTGGCAATTTTGCATGCTGAACCTTCTCAACAGAGAGAGAGGCTTCAACAGT 300  
OY 301 GTCTTTATGTAGATATATTAACCAAAAGAGGCTTCAAGAAAGTACAGATTTAGTC 360  
DB 301 GTCTTTATGTAGATATATTAACCAAAAGAGGCTTCAAGAAAGTACAGATTTAGTC 360  
OY 361 AACTTTGTAAGAGCTTATGAAATCATTTGTGCTTTTACAGCTTACACAGGTTGAGT 420  
DB 361 AACTTTGTAAGAGCTTATGAAATCATTTGTGCTTTTACAGCTTACACAGGTTGAGT 420  
OY 421 ATGCAACAGCTATATTTTGCAGAAAAAGAAAAATTAACCTCTCAACATCTAAAGATG 480  
DB 421 ATGCAACAGCTATATTTTGCAGAAAAAGAAAAATTAACCTCTCAACATCTAAAGATG 480  
OY 481 AAGTTTCTATCATCCAAAGTATGGGCTTACAGAAACCGTCCAAAAGCTTTTACAGAGTG 540  
DB 481 AAGTTTCTATCATCCAAAGTATGGGCTTACAGAAACCGTCCAAAAGCTTTTACAGAGTG 540  
OY 541 AACCCGAAATCTCTCTGAGAGAAACGCTCAGGTCCACACTCTTAACCTTGGAA 600  
DB 541 AACCCGAAATCTCTCTGAGAGAAACGCTCAGGTCCACACTCTTAACCTTGGAA 600  
OY 601 CTGTGAGAACTCTGAGAGCAAAAGGAGGATACAAACCTCAAAAACGCTCTCTACATTG 660  
DB 601 CTGTGAGAACTCTGAGAGCAAAAGGAGGATACAAACCTCAAAAACGCTCTCTACATTG 660  
OY 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGGCAACTTATGAGCTGTGGAG 720  
DB 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGGCAACTTATGAGCTGTGGAG 720  
OY 721 ATCAAGATTTGTTCAATATCAACCTTAAGAGAACAGGATGAATCAATGATTTGATTCG 780  
DB 721 ATCAAGATTTGTTCAATATCAACCTTAAGAGAACAGGATGAATCAATGATTTGATTCG 780  
OY 781 CAAAAAAGGCTGTTGAAATTTCTGAGAGGATGTAAACAAATATGACATCATCAAC 840  
DB 781 CAAAAAAGGCTGTTGAAATTTCTGAGAGGATGTAAACAAATATGACATCATCAAC 840  
OY 841 CCAATATTAATGATTTGAACACCACTGAGAGGCTGAGAGGATTCAGAGAAAGT 900  
DB 841 CCAATATTAATGATTTGAACACCACTGAGAGGCTGAGAGGATTCAGAGAAAGT 900  
OY 901 ATCAGGATGTTCTTTCAAACTTGCATGTGAGAGGATGGAGACAATATCTATGCCA 960  
DB 901 ATCAGGATGTTCTTTCAAACTTGCATGTGAGAGGATGGAGACAATATCTATGCCA 960

OY 961 GCTCATTTACAGCATGAGAACAGCACTTTATTTACTTACTTAAGAGAGATGATGAGAAA 1020  
DB 961 GCTCATTTACAGCATGAGAACAGCACTTTATTTACTTACTTAAGAGAGATGATGAGAAA 1020  
OY 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCCTGGCTTAGCAAGAGCCACATTAACAGAT 1080  
DB 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCCTGGCTTAGCAAGAGCCACATTAACAGAT 1080  
OY 1081 GGGGCGAAGTAAAGAAATGTAAATGATGAGGCGACTCCAGACACAAAAAAGGTAG 1140  
DB 1081 GGGGCGAAGTAAAGAAATGTAAATGATGAGGCGACTCCAGACACAAAAAAGGTAG 1140  
OY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATTAAGCAAACTGGCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATTAAGCAAACTGGCATGCT 1200  
OY 1201 CAGAGATCTAGAGATATCTAAGATGTTCTTGATATACACTAATATGACAGATTGAGA 1260  
DB 1201 CAGAGATCTAGAGATATCTAAGATGTTCTTGATATACACTAATATGACAGATTGAGA 1260  
OY 1261 AAGTTAATGAGTGGTTTCCAGAGTGAATGAACTGTAGGTTGATGATGATGATGATG 1320  
DB 1261 AAGTTAATGAGTGGTTTCCAGAGTGAATGAACTGTAGGTTGATGATGATGATGATG 1320  
OY 1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATTTGAGAGGTTCTAATATGAGTAGATG 1380  
DB 1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATTTGAGAGGTTCTAATATGAGTAGATG 1380  
OY 1381 AATATTTCTGTTCTTTCAGAGAAATATGACTTATGAGGCTGATGCTCATGAGGCTTTAA 1440  
DB 1381 AATATTTCTGTTCTTTCAGAGAAATATGACTTATGAGGCTGATGCTCATGAGGCTTTAA 1440  
OY 1441 TATGTAAAGTGAAGAGGTTCTACCTCAATATGATGAGAGATTAATTAAGCAAAATAT 1500  
DB 1441 TATGTAAAGTGAAGAGGTTCTACCTCAATATGATGAGAGATTAATTAAGCAAAATAT 1500  
OY 1501 TTGGGAAAACTTATGGAAGGCAAGCCTCCCACTTAAGCATATGTAAGTAAATC 1560  
DB 1501 TTGGGAAAACTTATGGAAGGCAAGCCTCCCACTTAAGCATATGTAAGTAAATC 1560  
OY 1561 TAAATTAAGGAGCAATTTGTTACTGAGCCACGATTAATTAAGAGAGGCTCCCTCAAAATA 1620  
DB 1561 TAAATTAAGGAGCAATTTGTTACTGAGCCACGATTAATTAAGAGAGGCTCCCTCAAAATA 1620  
OY 1621 AATTAAGGCTTAAGAGGAGCTACATCAGGCTTCAATCTGAGAGATTTTATCAAGAAAG 1680  
DB 1621 AATTAAGGCTTAAGAGGAGCTACATCAGGCTTCAATCTGAGAGATTTTATCAAGAAAG 1680  
OY 1681 CAGATTTGGCAGTTCAAAAGACTCTCTGAATGATAAATCAGAGAACTTAACCAAGAGAGC 1740  
DB 1681 CAGATTTGGCAGTTCAAAAGACTCTCTGAATGATAAATCAGAGAACTTAACCAAGAGAGC 1740  
OY 1741 AGAATGCTCAAGTATGATTTATTAATGATGCTCATGAGAAATTAAGCAAAAGGTGAT 1800  
DB 1741 AGAATGCTCAAGTATGATTTATTAATGATGCTCATGAGAAATTAAGCAAAAGGTGAT 1800  
OY 1801 CTATTCAGAAATGAGAAAAATCTTAACCAATTAAGATGATGATGAGAAAGAAATCTCTTCA 1860  
DB 1801 CTATTCAGAAATGAGAAAAATCTTAACCAATTAAGATGATGATGAGAAAGAAATCTCTTCA 1860  
OY 1861 AAAAGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGGAATCTGAATTAATATCC 1920  
DB 1861 AAAAGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGGAATCTGAATTAATATCC 1920  
OY 1921 ACAATTTCAAAAGCACTTAAGAAAGATAGGCTGAGAGAGAGATCTTCAACAGGATATTC 1980  
DB 1921 ACAATTTCAAAAGCACTTAAGAAAGATAGGCTGAGAGAGAGATCTTCAACAGGATATTC 1980  
OY 1981 ATGGGCTTGAACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
DB 1981 ATGGGCTTGAACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
OY 2041 TTGATAGCTTCTTGAAGAGTAAAGAGATTAAGAAAAAAGTACCAACCAATATGCTCA 2100

Db 2041 TTGATGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAGTACAAACCAATGCGACATCA 2100  
 QY 2101 GGCACAGCAGAAACCTACACACTCATGGAAGTAAGAACTGCACTGGAGCCAGAGAGA 2160  
 Db 2101 GGCACAGCAGAAACCTACACACTCATGGAAGTAAGAACTGCACTGGAGCCAGAGAGA 2160  
 QY 2161 GTACACAGCCAAATGAAGACAGACAGTAAGAGACATGAGTACTGATCTTCCAGAGCTGA 2220  
 Db 2161 GTACACAGCCAAATGAAGACAGACAGTAAGAGACATGAGTACTGATCTTCCAGAGCTGA 2220  
 QY 2221 AGTTAAACAATGACACCTGGTCTTTTACTAAGTGTTCAAATACAGGAACCTTAAAGAT 2280  
 Db 2221 AGTTAAACAATGACACCTGGTCTTTTACTAAGTGTTCAAATACAGGAACCTTAAAGAT 2280  
 QY 2281 TTGTCAATCCTACCTTCCAGAGAGAAAAAGAGAAACTGAAACAGTTAAAGTGT 2340  
 Db 2281 TTGTCAATCCTTACCTTCCAGAGAGAAAAAGAGAAACTGAAACAGTTAAAGTGT 2340  
 QY 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAAGTGAAGAGGTTTGCACACTG 2400  
 Db 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAAGTGAAGAGGTTTGCACACTG 2400  
 QY 2401 AAAGATCTGTAGAGTACGAGTATTTTCATTTGCTGACTGATTTATGGACCTGAG 2460  
 Db 2401 AAAGATCTGTAGAGTACGAGTATTTTCATTTGCTGACTGATTTATGGACCTGAG 2460  
 QY 2461 AAAGTATCTGCTAGTGAAGAGTGAAGTACGCTAGGGAAGGCAAAAACGAACCAATTAAT 2520  
 Db 2461 AAAGTATCTGCTAGTGAAGAGTGAAGTACGCTAGGGAAGGCAAAAACGAACCAATTAAT 2520  
 QY 2521 GTGTGAGTCACTGTGAGCATTTTGAAGAACCCCAAGGACTAATTCATGTTGTTCCAAAG 2580  
 Db 2521 GTGTGAGTCACTGTGAGCATTTTGAAGAACCCCAAGGACTAATTCATGTTGTTCCAAAG 2580  
 QY 2581 ATATATGAAATGACACAGAAAGCTTTAATATCCATTGGACATGAAGTTAACACAGTC 2640  
 Db 2581 ATATATGAAATGACACAGAAAGCTTTAATATCCATTGGACATGAAGTTAACACAGTC 2640  
 QY 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAAGTGTGCTCAATTTTTCAGAAATTCAT 2700  
 Db 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAAGTGTGCTCAATTTTTCAGAAATTCAT 2700  
 QY 2701 TCAAGGTTTCAAGGCCCAAGTCATTTGCTGTGTTTCMAATCCAGAAATGCAGAAAGAG 2760  
 Db 2701 TCAAGGTTTCAAGGCCCAAGTCATTTGCTGTGTTTCMAATCCAGAAATGCAGAAAGAG 2760  
 QY 2761 AATGTGCAACATTTCTGCCCACCTGTGGCTCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
 Db 2761 AATGTGCAACATTTCTGCCCACCTGTGGCTCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
 QY 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGATGAGTCAATATCAGGCTTAC 2880  
 Db 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGATGAGTCAATATCAGGCTTAC 2880  
 QY 2881 AGACAGTTAATATCACTGCAAGGCTTCTGTGTGTGTGTCAGAAAGATTAAGCCAGTTGATA 2940  
 Db 2881 AGACAGTTAATATCACTGCAAGGCTTCTGTGTGTGTGTCAGAAAGATTAAGCCAGTTGATA 2940  
 QY 2941 ATGCCCAATGTAGTATCAAGAGAGGCTTAGGTTTGTCTATATCTCACTTCAGAGGCA 3000  
 Db 2941 ATGCCCAATGTAGTATCAAGAGAGGCTTAGGTTTGTCTATATCTCACTTCAGAGGCA 3000  
 QY 3001 AGCAAACTGACATTTACTCCCAATTAACATGAGCTTTTACAAAACCCATATCGTATAC 3060  
 Db 3001 AGCAAACTGACATTTACTCCCAATTAACATGAGCTTTTACAAAACCCATATCGTATAC 3060  
 QY 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTAGAAAAAATCTCTAGAGG 3120  
 Db 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTAGAAAAAATCTCTAGAGG 3120  
 QY 3121 AAAACTTTGAGAACATTTCAATGTCACTGAAAGAGAAATGGAAATGAGACATTTCCAA 3180  
 Db 3121 AAAACTTTGAGAACATTTCAATGTCACTGAAAGAGAAATGGAAATGAGACATTTCCAA 3180

Db 3121 AAAACTTTGAGAACATTTCAATGTCACTGAAAGAGAAATGGAAATGAGACATTTCCAA 3180  
 QY 3181 GTACAGTGAACACATTTAGCCGTAATTAATTAATGAGAAATGTTTTTAAAGAACCCAGCT 3240  
 Db 3181 GTACAGTGAACACATTTAGCCGTAATTAATTAATGAGAAATGTTTTTAAAGAACCCAGCT 3240  
 QY 3241 CAAGCAATATTAAATGAAGTGAAGTCCAGTACTAATGAAGTGGGCTCCAGATTTAATGA 3300  
 Db 3241 CAAGCAATATTAAATGAAGTGAAGTCCAGTACTAATGAAGTGGGCTCCAGATTTAATGA 3300  
 QY 3301 TAGGTTCCAGTATGAAGAAACATTCAGCAGAACTAGTATAGAAACAGAGGCCAAATTTGA 3360  
 Db 3301 TAGGTTCCAGTATGAAGAAACATTCAGCAGAACTAGTATAGAAACAGAGGCCAAATTTGA 3360  
 QY 3361 ATGCTATGCTTAGATTGAGGTTTGGACCTGAGTCTCTTAACAAAGTCTCTCGAA 3420  
 Db 3361 ATGCTATGCTTAGATTGAGGTTTGGACCTGAGTCTCTCTTAACAAAGTCTCTCGAA 3420  
 QY 3421 GTAAATTTGAAGCATCTGAAATTAAGAAAGCAGAAATATGAAGATAGTTTCAAGCTGTTA 3480  
 Db 3421 GTAAATTTGAAGCATCTGAAATTAAGAAAGCAGAAATATGAAGATAGTTTCAAGCTGTTA 3480  
 QY 3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAATTAACAGCCTATGGAGATAGTC 3540  
 Db 3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAATTAACAGCCTATGGAGATAGTC 3540  
 QY 3541 ATGCATCTCAGGTTTGTTCGAGACACCTGATACCTGTTAGTGAATGGAATTAAGG 3600  
 Db 3541 ATGCATCTCAGGTTTGTTCGAGACACCTGATACCTGTTAGTGAATGGAATTAAGG 3600  
 QY 3601 AAGATACTAGTTTGTTCGAAATTAAGAAATGACATTAAGAAAGTCTGCTGTTTACAAAGCG 3660  
 Db 3601 AAGATACTAGTTTGTTCGAAATTAAGAAATGACATTAAGAAAGTCTGCTGTTTACAAAGCG 3660  
 QY 3661 TCCAGAAAGACAGCTTAGCAGAGATCTAGCCCTTTACACCATACACATTTGGCTCAGG 3720  
 Db 3661 TCCAGAAAGACAGCTTAGCAGAGATCTAGCCCTTTACACCATACACATTTGGCTCAGG 3720  
 QY 3721 GTTACCGAAGAGGGGCCCAAGAAATTAAGATAGAGTCCAGAGAAAGCTATCTAGGAGAGG 3780  
 Db 3721 GTTACCGAAGAGGGGCCCAAGAAATTAAGATAGAGTCCAGAGAAAGCTATCTAGGAGAGG 3780  
 QY 3781 AAGAGCTTCCCTGCTCCACACACTGTTATTTGGTAAAGTAAACAATATACCTTCTCACT 3840  
 Db 3781 AAGAGCTTCCCTGCTCCACACACTGTTATTTGGTAAAGTAAACAATATACCTTCTCACT 3840  
 QY 3841 CTACTAGCATTAACACCGTGTGCTACCGAGTGTGCTTAAGAAACAGAGAGAAATTTAT 3900  
 Db 3841 CTACTAGCATTAACACCGTGTGCTACCGAGTGTGCTTAAGAAACAGAGAGAAATTTAT 3900  
 QY 3901 TATCATTTGAAGATTAAGTAAATGACTGACGTAACAGGTAATATTTGGCAAGGCACTTC 3960  
 Db 3901 TATCATTTGAAGATTAAGTAAATGACTGACGTAACAGGTAATATTTGGCAAGGCACTTC 3960  
 QY 3961 AGGAACATCACTTAAGAGAGAAACAAATGTTCTGCTAGTGTGTTTCTTCAAGTGA 4020  
 Db 3961 AGGAACATCACTTAAGAGAGAAACAAATGTTCTGCTAGTGTGTTTCTTCAAGTGA 4020  
 QY 4021 GTGAATTTGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
 Db 4021 GTGAATTTGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
 QY 4081 CCAAAACAATGAGGCACTCACTGGAAGGCAAGGAGATTTGGTCTGAGTGCAGAAAGATTTGG 4140  
 Db 4081 CCAAAACAATGAGGCACTCACTGGAAGGCAAGGAGATTTGGTCTGAGTGCAGAAAGATTTGG 4140  
 QY 4141 TTTTCAGATGATGAAGAAAGAGAAAGGAGGCTTGAAGAAAGAAATTAATGAAGAGCAAGAGA 4200  
 Db 4141 TTTTCAGATGATGAAGAAAGAGAAAGGAGGCTTGAAGAAAGAAATTAATGAAGAGCAAGAGA 4200  
 QY 4201 TGGATTTCAAACTTAGTGAAGAGAGATCTGGGTGTGAGAGTGAAGAGAGAGAGAGAGAG 4260  
 Db 4201 TGGATTTCAAACTTAGTGAAGAGAGATCTGGGTGTGAGAGTGAAGAGAGAGAGAGAGAGAG 4260





Qy	301	GTCTTTTATGTAGAATGATATATACCAAAAGGAGCCCTACAGAAAGTACGAGATTTACTC	360
Db	301	GTCTTTTATGTAGAATGATATATACCAAAAGGAGCCCTACAGAAAGTACGAGATTTACTC	360
Qy	361	AACTTGTTGAACGCTATTGAAATCATTTTGGCTTTTCAGCTTGACACAGGTTTGGAGT	420
Db	361	AACTTGTTGAACGCTATTGAAATCATTTTGGCTTTTCAGCTTGACACAGGTTTGGAGT	420
Qy	421	ATGCAAAACGCTATATATTTTGCAAAAAGAAATATACCTCCGTACATCTTAANAATG	480
Db	421	ATGCAAAACGCTATATATTTTGCAAAAAGAAATATACCTCCGTACATCTTAANAATG	480
Qy	481	AAGTTTCTATCATCCAAAGATAGGCTCACGAACCGTGCCAAAAGCTTCTACAGATG	540
Db	481	AAGTTTCTATCATCCAAAGATAGGCTCACGAACCGTGCCAAAAGCTTCTACAGATG	540
Qy	541	AACCCGAAAATCTTCCTTGACGAGAACACAGTCTAGTGCCAACTCTTAACCTGGAA	600
Db	541	AACCCGAAAATCTTCCTTGACGAGAACACAGTCTAGTGCCAACTCTTAACCTGGAA	600
Qy	601	CTGTGAGAACTCTGAGGACAAAGACGCGGATACAACCTCAAAAAGACGCTGTCTACATG	660
Db	601	CTGTGAGAACTCTGAGGACAAAGACGCGGATACAACCTCAAAAAGACGCTGTCTACATG	660
Qy	661	AATTGGGATCTGATCTTCTTGAAATACCGTTAATAAGCAACTATTGCACTGTGGAG	720
Db	661	AATTGGGATCTGATCTTCTTGAAATACCGTTAATAAGCAACTATTGCACTGTGGAG	720
Qy	721	ATCAGAATTTGTTTACAAATCACCCCTCAAGGACACCGAGATGAATTCAGTTGGATTGTG	780
Db	721	ATCAGAATTTGTTTACAAATCACCCCTCAAGGACACCGAGATGAATTCAGTTGGATTGTG	780
Qy	781	CAAAAAGCGCTGTGTGAATTTTCTGAGCGGATGTAAACAATCTGACATCATCAAC	840
Db	781	CAAAAAGCGCTGTGTGAATTTTCTGAGCGGATGTAAACAATCTGACATCATCAAC	840
Qy	841	CCAGTAATTAATGATTTGGAACACCACTGAGAAAGCGTGCAGTGAAGGCGATCCAGAAAAGT	900
Db	841	CCAGTAATTAATGATTTGGAACACCACTGAGAAAGCGTGCAGTGAAGGCGATCCAGAAAAGT	900
Qy	901	ATCAGGGATGTTCTGTTCAAACTTGATGTGAGCCCATGTGGCACAAAATCTCATGCGCA	960
Db	901	ATCAGGGATGTTCTGTTCAAACTTGATGTGAGCCCATGTGGCACAAAATCTCATGCGCA	960
Qy	961	GCTCAATTACAGATGAGAACAGCAAGCATTTATTAATCACTAAAGACAGAAATGATGTAGAAA	1020
Db	961	GCTCAATTACAGATGAGAACAGCAAGCATTTATTAATCACTAAAGACAGAAATGATGTAGAAA	1020
Qy	1021	AGGCTGAATTTCTGTATATAAAGCAAAACGCGTGTGCAAGAGAGCCACATTAACGAT	1080
Db	1021	AGGCTGAATTTCTGTATATAAAGCAAAACGCGTGTGCAAGAGAGCCACATTAACGAT	1080
Qy	1081	GGGCTGGAAGTAAAGAAACATGTATAGTAGCGGACTCCAGCACAGAAAAAAGGTAG	1140
Db	1081	GGGCTGGAAGTAAAGAAACATGTATAGTAGCGGACTCCAGCACAGAAAAAAGGTAG	1140
Qy	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGGAATTAACACAGAAATGCGCATGCT	1200
Db	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGGAATTAACACAGAAATGCGCATGCT	1200
Qy	1201	CAGAGAAATCCTAGAGATCTACTGAAAGTGTCCCTGGATTAACACTTAATTAAGCAGATTCAGA	1260
Db	1201	CAGAGAAATCCTAGAGATCTACTGAAAGTGTCCCTGGATTAACACTTAATTAAGCAGATTCAGA	1260
Qy	1261	AACTTAATGAGTGGTTTTCCAGAAAGTATGAACCTGTTAGTTCGATGACTACATGATG	1320
Db	1261	AACTTAATGAGTGGTTTTCCAGAAAGTATGAACCTGTTAGTTCGATGACTACATGATG	1320
Qy	1321	GGGAGTGTGAATCAAAATCCCAAAGTAGCTGTATGTGGAGAGCTTCAATGAAGGTAGATG	1380
Db	1321	GGGAGTGTGAATCAAAATCCCAAAGTAGCTGTATGTGGAGAGCTTCAATGAAGGTAGATG	1380

QY	1381	AATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCTCTATAGGCTTTAA	1440
Db	1381	AATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCTCTATAGGCTTTAA	1440
QY	1441	TATGTAAAGGAAGAGCTTCACCTCCAAATCAGTAGACAGTAATTTGACACAAAATAT	1500
Db	1441	TATGTAAAGGAAGAGCTTCACCTCCAAATCAGTAGACAGTAATTTGACACAAAATAT	1500
QY	1501	TTGGGAAAACCTATGCGAAGAGCAACCTCCCAACTTTAAGCATTTAATCTGAAAATC	1560
Db	1501	TTGGGAAAACCTATGCGAAGAGCAACCTCCCAACTTTAAGCATTTAATCTGAAAATC	1560
QY	1561	TAAATTATAGAGCACTTTGTACTGAGCCACAGATTAATACAGACGTCCTCCACAAATA	1620
Db	1561	TAAATTATAGAGCACTTTGTACTGAGCCACAGATTAATACAGACGTCCTCCACAAATA	1620
QY	1621	AATTAAACCGTAAAGGAGACCTACATAGGCTTCATCTGAGAGTTTATVCAAGAAAG	1680
Db	1621	AATTAAACCGTAAAGGAGACCTACATAGGCTTCATCTGAGAGTTTATVCAAGAAAG	1680
QY	1681	CAGATTGTGGCAGTTCAAAAGAGCTCCCTGAAATGATAAATCAGAGAACTAACCAACGAGC	1740
Db	1681	CAGATTGTGGCAGTTCAAAAGAGCTCCCTGAAATGATAAATCAGAGAACTAACCAACGAGC	1740
QY	1741	AGATGTGTCAAGTGTATGATTAATTAATTAAGTGTGTCATGAGATTAACCAAGAGTAT	1800
Db	1741	AGATGTGTCAAGTGTATGATTAATTAATTAAGTGTGTCATGAGATTAACCAAGAGTAT	1800
QY	1801	CTATTCAAAATAGAGAAAATCCTTAACCCAAATAGAAATCCTGAAAAGAAATCTCTTTCA	1860
Db	1801	CTATTCAAAATAGAGAAAATCCTTAACCCAAATAGAAATCCTGAAAAGAAATCTCTTTCA	1860
QY	1861	AAACGAAGCTGGAACCTTAAGCAGCAGTATTAAGCAATATGGAATCGAATTAATATTC	1920
Db	1861	AAACGAAGCTGGAACCTTAAGCAGCAGTATTAAGCAATATGGAATCGAATTAATATTC	1920
QY	1921	ACAATTCAAAAGCACTTAAGAAAGATAGGCTGAGAGGAAGTCTTACCAAGCATATTC	1980
Db	1921	ACAATTCAAAAGCACTTAAGAAAGATAGGCTGAGAGGAAGTCTTACCAAGCATATTC	1980
QY	1981	ATGGGCTTGACACTGTAAGTGTGAATTAACCCCAATTTGTAAGTGAATTTGCAAA	2040
Db	1981	ATGGGCTTGACACTGTAAGTGTGAATTAACCCCAATTTGTAAGTGAATTTGCAAA	2040
QY	2041	TTGATAGTGTTCACAGAGTGAAGAGATTAAGAAAAAAAGTATCAACCAATGGCAGTCA	2100
Db	2041	TTGATAGTGTTCACAGAGTGAAGAGATTAAGAAAAAAAGTATCAACCAATGGCAGTCA	2100
QY	2101	GGCAGCAGAGAAACCTTACACTCATGTGAGAGTAAAGAAACCTGCAACTGAGGCCAAGAGA	2160
Db	2101	GGCAGCAGAGAAACCTTACACTCATGTGAGAGTAAAGAAACCTGCAACTGAGGCCAAGAGA	2160
QY	2161	GTAACAAGGCCAAATGAAGACAGCAAGTAAAGACATGACAGTAACTTCCAGAGCTGA	2220
Db	2161	GTAACAAGGCCAAATGAAGACAGCAAGTAAAGACATGACAGTAACTTCCCAAGCTGA	2220
QY	2221	AGTTAAACAAATGACACCTGGTCTTTTACTAAGTGTAAATATACAGTGAATTAAGAAAT	2280
Db	2221	AGTTAAACAAATGACACCTGGTCTTTTACTAAGTGTAAATATACAGTGAATTAAGAAAT	2280
QY	2281	TTGTCAATCTTAGCCTTCCAAAGAGAGAAAAAGAGAAAACTAGAAACAGTTAAAGTGT	2340
Db	2281	TTGTCAATCTTAGCCTTCCAAAGAGAGAAAAAGAGAAAACTAGAAACAGTTAAAGTGT	2340
QY	2341	CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGGTTTTGCAAACTG	2400
Db	2341	CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGGTTTTGCAAACTG	2400
QY	2401	AAACATCTGTAGAGAGTAGACAGTATTTCTATTTGTAAGTGTACTGATTAATGCGACTAGG	2460
Db	2401	AAACATCTGTAGAGAGTAGACAGTATTTCTATTTGTAAGTGTACTGATTAATGCGACTAGG	2460
QY	2461	AAAGTATCTGCTTACTGGAAGTTTGGCACTGTAAGGAGGCAAAAAACGAGACCAATTAAT	2520



```
Db 2461 ||||| AAAAGTATCTCGTTACTGGAAAGTTAGCACTTAGGGAAAGCAAAAACAACCAATTAAT 2520
Qy 2521 GTGTAGTCAGTGTGTCAGCATTTGAAACCCCAAGGAGCTAATTCATGTTGTTCCAAAG 2580
Db 2521 GTGTAGTCAGTGTGTCAGCATTTGAAACCCCAAGGAGCTAATTCATGTTGTTCCAAAG 2580
Qy 2581 ATAAATGAATGACACAGAAAGGCTTTAAGTATCCATTGGACATGAACTTAACCAAGTTC 2640
Db 2581 ATAAATGAATGACACAGAAAGGCTTTAAGTATCCATTGGACATGAACTTAACCAAGTTC 2640
Qy 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCCAATTAAT 2700
Db 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCCAATTAAT 2700
Qy 2701 TCAAGGTTTCAAGGGCCAGTATCTGCTGTTTTCAAATCCAGAAATGAGAAAGG 2760
Db 2701 TCAAGGTTTCAAGGGCCAGTATCTGCTGTTTTCAAATCCAGAAATGAGAAAGG 2760
Qy 2761 AATGTGCAACATTTCTCTGCCACTCTGGGTCCTTAAAGAAACAAGTCCAAAGTCACTT 2820
Db 2761 AATGTGCAACATTTCTCTGCCACTCTGGGTCCTTAAAGAAACAAGTCCAAAGTCACTT 2820
Qy 2821 TTGAATGTGAACAAAGAGAGAAAATCAAGGAAGATGAGTCTAATATCAAGCTGTAC 2880
Db 2821 TTGAATGTGAACAAAGAGAGAAAATCAAGGAAGATGAGTCTAATATCAAGCTGTAC 2880
Qy 2881 AGACGTTAATATCTGCTGAGGCTTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2940
Db 2881 AGACGTTAATATCTGCTGAGGCTTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2940
Qy 2941 ATGCCAAATGTATGATCAAGAGAGGCTAGGTTTGTCTATCATCTGAGTTGACAGAGCA 3000
Db 2941 ATGCCAAATGTATGATCAAGAGAGGCTAGGTTTGTCTATCATCTGAGTTGACAGAGCA 3000
Qy 3001 ACGAAACGAGCTCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATCTCTATAC 3060
Db 3001 ACGAAACGAGCTCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATCTCTATAC 3060
Qy 3061 CACCACTTTTCCCTCAAGTCATTTGTTAAACATTAATGAAGAAATGCTGAGAGG 3120
Db 3061 CACCACTTTTCCCTCAAGTCATTTGTTAAACATTAATGAAGAAATGCTGAGAGG 3120
Qy 3121 AAACTTTGAGGAACATTCATGTCACCTGGAAGAAATGGAATGAGAATGAGATTCCTCA 3180
Db 3121 AAACTTTGAGGAACATTCATGTCACCTGGAAGAAATGGAATGAGAATGAGATTCCTCA 3180
Qy 3181 GTACAGTACGACAAATTAAGCCGTAATTAACATTAGAGAAATGTTTTTAAAGAGCCAGCT 3240
Db 3181 GTACAGTACGACAAATTAAGCCGTAATTAACATTAGAGAAATGTTTTTAAAGAGCCAGCT 3240
Qy 3241 CAAGCAATTAATTAAGTAGTCCAGTACTAATGAAGAGGCGCCAGTATTAATGAAG 3300
Db 3241 CAAGCAATTAATTAAGTAGTCCAGTACTAATGAAGAGGCGCCAGTATTAATGAAG 3300
Qy 3301 TAGGTTCCAGTATGAAAAACATTCGAAGCAAGTAGTGAAGAAAGAGAGCCCAAAATGGA 3360
Db 3301 TAGGTTCCAGTATGAAAAACATTCGAAGCAAGTAGTGAAGAAAGAGAGCCCAAAATGGA 3360
Qy 3361 ATGCTATGCTTAGATTAAGGGGTTTTGCAACCTGAGCTTATTAACAAAGCTTCTCTGAA 3420
Db 3361 ATGCTATGCTTAGATTAAGGGGTTTTGCAACCTGAGCTTATTAACAAAGCTTCTCTGAA 3420
Qy 3421 GTAATTTAGAGATCTGAAATTAAGAAAGCAAGATTAAGAAAGTATGAGATGAGATGTA 3480
Db 3421 GTAATTTAGAGATCTGAAATTAAGAAAGCAAGATTAAGAAAGTATGAGATGAGATGTA 3480
Qy 3481 ATACAGATTTCTCCATATCTGATTTAGATTAACATTAGAACAGGCTTATGGAAGTATGTC 3540
Db 3481 ATACAGATTTCTCCATATCTGATTTAGATTAACATTAGAACAGGCTTATGGAAGTATGTC 3540
Qy 3541 ATGCAATCAGGTTTGTCTGAGACCTGATGACCTTTGATGATGATGCTGAATTAAGG 3600
Db 3541 ATGCAATCAGGTTTGTCTGAGACCTGATGACCTTTGATGATGATGCTGAATTAAGG 3600
```

---

```
Db 3541 ATGCAATCAGGTTTGTCTGAGACCTGATGACCTGTTAGATGATGCTGAATTAAGG 3600
Qy 3601 AAGATAGTATGTTTGTCTGAAAAATGACATTAAGAAAGTTCTGCTGTTTATACCAAGCG 3660
Db 3601 AAGATAGTATGTTTGTCTGAAAAATGACATTAAGAAAGTTCTGCTGTTTATACCAAGCG 3660
Qy 3661 TCCAGAAAGAGAGGCTTAGCAGAGTCCTAGCCCTTTCACCCATACACATTTGGCTCAGG 3720
Db 3661 TCCAGAAAGAGAGGCTTAGCAGAGTCCTAGCCCTTTCACCCATACACATTTGGCTCAGG 3720
Qy 3721 GTTACCAGAAAGAGGGGCCAAGAAATTAAGATCCTCAGAAAGAACTTATCTAGTAGATG 3780
Db 3721 GTTACCAGAAAGAGGGGCCAAGAAATTAAGATCCTCAGAAAGAACTTATCTAGTAGATG 3780
Qy 3781 AAGACCTTCCCTGCTTCCAAACCTTGTAATTTGTAAGTAACATATACCTTCAGT 3840
Db 3781 AAGACCTTCCCTGCTTCCAAACCTTGTAATTTGTAAGTAACATATACCTTCAGT 3840
Qy 3841 CTACTAGGCAATAGCACCCGTTGCTACCGAGTGTCTGTAAAGAACACAGAGAGAAATTAAT 3900
Db 3841 CTACTAGGCAATAGCACCCGTTGCTACCGAGTGTCTGTAAAGAACACAGAGAGAAATTAAT 3900
Qy 3901 TATCATTTGAGATATGCTTAATTAATGATGCTGATTAACAGATTAATTTGGCAAGCATCTC 3960
Db 3901 TATCATTTGAGATATGCTTAATTAATGATGCTGATTAACAGATTAATTTGGCAAGCATCTC 3960
Qy 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTGCTAGAGTGTGTTTCTTCCAGAGTGA 4020
Db 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTGCTAGAGTGTGTTTCTTCCAGAGTGA 4020
Qy 4021 GTGAATTTGGAAGACTTGAAGTGAACAAACACCCAGAGATCTTCTGATTTGTTGTTCT 4080
Db 4021 GTGAATTTGGAAGACTTGAAGTGAACAAACACCCAGAGATCTTCTGATTTGTTGTTCT 4080
Qy 4081 CCAAAACAATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4140
Db 4081 CCAAAACAATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4140
Qy 4141 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4200
Db 4141 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4200
Qy 4201 TGGATTTCAAACCTTAGTGAAGCAGATCTGGGTTGAGAGTGAAGCAAGCGCTCTGAAAG 4260
Db 4201 TGGATTTCAAACCTTAGTGAAGCAGATCTGGGTTGAGAGTGAAGCAAGCGCTCTGAAAG 4260
Qy 4261 ACTGCTCAGGCTATCTCTCTCAGAGTGAATTTAACCACTCAGCAGAGAGGATFACATGC 4320
Db 4261 ACTGCTCAGGCTATCTCTCTCAGAGTGAATTTAACCACTCAGCAGAGAGGATFACATGC 4320
Qy 4321 AACATTAACCTGATTAAGGCTCCAGAGAAATGAGTGAAGTGAAGCTGTTAGAAAGC 4380
Db 4321 AACATTAACCTGATTAAGGCTCCAGAGAAATGAGTGAAGTGAAGCTGTTAGAAAGC 4380
Qy 4381 ATGGAGGCGAGCTTCTTAACAGCTACCTTCATGATTAAGTACCTTCCTGCTTGAAG 4440
Db 4381 ATGGAGGCGAGCTTCTTAACAGCTACCTTCATGATTAAGTACCTTCCTGCTTGAAG 4440
Qy 4441 ACCTCGGAATTCAGACAAAGCAGATCAGAAAAAGCAGTATTAACCTTACAGAAAAAGTA 4500
Db 4441 ACCTCGGAATTCAGACAAAGCAGATCAGAAAAAGCAGTATTAACCTTACAGAAAAAGTA 4500
Qy 4501 GTGAATTAACCTTATAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTTGAAGTCTG 4560
Db 4501 GTGAATTAACCTTATAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTTGAAGTCTG 4560
Qy 4561 CAGATAGTCTTACCAAGTAAAAATAAGAACAGAGAGTGAAGAGTCAATCCCTTTAAAT 4620
Db 4561 CAGATAGTCTTACCAAGTAAAAATAAGAACAGAGAGTGAAGAGTCAATCCCTTTAAAT 4620
Qy 4621 GCCCATCAATTAAGATATAGGTTGATGATGATGATGATGATGATGATGATGATGATGATG 4680
Db 4621 GCCCATCAATTAAGATATAGGTTGATGATGATGATGATGATGATGATGATGATGATGATG 4680
```

OY 4681 ACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGAGAGACAACAGCTGAGAG 4740  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4681 ACTACCATCTCAAGAGAGAGCTCATTAAGGTTGTTGATGAGAGAGACAACAGCTGAGAG 4740  
 OY 4741 AGTCTGGGCCACACGATTTTACGGAACATCTTACTTCCAGAGCAAGATCTAGAGGAA 4800  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4741 AGTCTGGGCCACACGATTTTACGGAACATCTTACTTCCAGAGCAAGATCTAGAGGAA 4800  
 OY 4801 CCCCTTACCTGSAATCTGSAATCAGCCTCTCTCTGATGAGCCCTGSAATCTGATCTCTG 4860  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4801 CCCCTTACCTGSAATCTGSAATCAGCCTCTCTCTGATGAGCCCTGSAATCTGATCTCTG 4860  
 OY 4861 AAGACAGAGCCCCAGAGATCTGCTGTTGGCAACATACATCTTCAACCTCTGATTTGA 4920  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4861 AAGACAGAGCCCCAGAGATCTGCTGTTGGCAACATACATCTTCAACCTCTGATTTGA 4920  
 OY 4921 AAGTCCCAATTTGAAAGTTGACAGATCTGCCAGACTCCAGCTGCTGCTCATCTACTG 4980  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4921 AAGTCCCAATTTGAAAGTTGACAGATCTGCCAGACTCCAGCTGCTGCTCATCTACTG 4980  
 OY 4981 ATACTGCTGGGTAAATGCAATGGAAGAAGTGCACAGGAGAGAGCAAGCAATTTGACAG 5040  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 4981 ATACTGCTGGGTAAATGCAATGGAAGAAGTGCACAGGAGAGAGCAAGCAATTTGACAG 5040  
 OY 5041 CTTCACAGAGAAAGGCTCAACAAAAGAAATGTCATGCTGCTGCTGAGCCCTGACCCAGAG 5100  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5041 CTTCACAGAGAAAGGCTCAACAAAAGAAATGTCATGCTGCTGAGCCCTGACCCAGAG 5100  
 OY 5101 AATTTATGCTCGGTACAGATTTTCCAGAAAACACACATCATCTTAACTAATCTAATTA 5160  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5101 AATTTATGCTCGGTACAGATTTTCCAGAAAACACACATCATCTTAACTAATCTAATTA 5160  
 OY 5161 CTGAAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTGTTGTTGTTGTTGTTG 5220  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5161 CTGAAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTGTTGTTGTTGTTGTTG 5220  
 OY 5221 TGAATAATTTTCTAGGAATTTGCGGAGAGAAAATGGTAGTTAGCTAATTTTCTGGTGACC 5280  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5221 TGAATAATTTTCTAGGAATTTGCGGAGAGAAAATGGTAGTTAGCTAATTTTCTGGTGACC 5280  
 OY 5281 AGCTCTTTAAAGAAAGAAAATGCTGAAATGAGCATGATTTTGAAGTGAAGAGAGATGAGG 5340  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5281 AGCTCTTTAAAGAAAGAAAATGCTGAAATGAGCATGATTTTGAAGTGAAGAGAGATGAGG 5340  
 OY 5341 TCAATGGAAGAAACCAACCAAGTTCACAAAGCAGCAAGAGATCCAGAGACAGAAAGATCT 5400  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5341 TCAATGGAAGAAACCAACCAAGTTCACAAAGCAGCAAGAGATCCAGAGACAGAAAGATCT 5400  
 OY 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTCACCAACATGCCCACAGATCAACTGG 5460  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTCACCAACATGCCCACAGATCAACTGG 5460  
 OY 5461 AATGAGTGTGACAGCTGTGTGTGCTCTGAGGAGAGAGCTTCAATCACTTCACTCCCTG 5520  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5461 AATGAGTGTGACAGCTGTGTGTGCTCTGAGGAGAGAGCTTCAATCACTTCACTCCCTG 5520  
 OY 5461 AATGAGTGTGACAGCTGTGTGTGCTCTGAGGAGAGAGCTTCAATCACTTCACTCCCTG 5520  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5461 AATGAGTGTGACAGCTGTGTGTGCTCTGAGGAGAGAGCTTCAATCACTTCACTCCCTG 5520  
 OY 5521 GCACAGGTGTCCACCAATTTGTTGCTGTCAGAGCAGATGCTGAGAGAGAGATGAGT 5580  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5521 GCACAGGTGTCCACCAATTTGTTGCTGTCAGAGCAGATGCTGAGAGAGAGATGAGT 5580  
 OY 5581 TCATGCAATTTGGGAGATGTTGAGGACACTGTTGTTGAGCCGAGAGTGGGTTGGACA 5640  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5581 TCATGCAATTTGGGAGATGTTGAGGACACTGTTGTTGAGCCGAGAGTGGGTTGGACA 5640  
 OY 5641 GTGTAGACTCTTCAGATGCTGAGAGAGTGCACACTCTGATATCCCAAGATCCCGACA 5700  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5641 GTGTAGACTCTTCAGATGCTGAGAGAGTGCACACTCTGATATCCCAAGATCCCGACA 5700  
 OY 5701 GCCACTACTGA 5711  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 5701 GCCACTACTGA 5711

RESULT 8  
 AR184044  
 LOCUS AR184044 5712 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 47 from patent US 6342483.  
 ACCESSION AR184044  
 VERSION AR184044.1 GI:20228013  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unclassified.  
 REFERENCE  
 1 (bases 1 to 5712)  
 HOLZ, J.T., Jensen, R.A., Page, D.L., Obermiller, P.S.,  
 Robinson-Benion, C.L., and Thompson, M.E.  
 TITLE Method for detection and treatment of breast cancer  
 JOURNAL Patent: US 6342483-A 47 29-JAN-2002;  
 FEATURES  
 source  
 1..5712  
 Location/Qualifiers  
 BASE COUNT 1956 a 1099 c 1274 g 1383 t  
 ORIGIN  
 Query Match 99.98; Score 5707.8; DB 6; Length 5712;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY- 1 AGCTGCGTGAAGACTTCCTGGACCCCGCAGCAGGCGTGGGGTTTCCAGATACTGGGCC 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 1 AGCTGCGTGAAGACTTCCTGGACCCCGCAGCAGGCGTGGGGTTTCCAGATACTGGGCC 60  
 OY 61 CCTGCGCTCAGAGAGGCGCTTCAACCTCTGCTGCTGGGTAAGATTGATTTGGAACAGAGAAA 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 61 CCTGCGCTCAGAGAGGCGCTTCAACCTCTGCTGCTGGGTAAGATTGATTTGGAACAGAGAAA 120  
 OY 121 TGGATTTATGCTGCTGCTGCGGTTGAAGAGTACAAAATGCTAATTAATGATGACAGAAA 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 121 TGGATTTATGCTGCTGCTGCGGTTGAAGAGTACAAAATGCTAATTAATGATGACAGAAA 180  
 OY 181 TCTTAGAGTGTCCCATCTGCTGAGAGTTGATCAAGAACTGCTCCACAAAGCTGAGC 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 181 TCTTAGAGTGTCCCATCTGCTGAGAGTTGATCAAGAACTGCTCCACAAAGCTGAGC 240  
 OY 241 ACATATTTTGAATTTTTCATGCTGAACTTCTTCAACCAAGAAAGAGGCTTTCACAGT 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 241 ACATATTTTGAATTTTTCATGCTGAACTTCTTCAACCAAGAAAGAGGCTTTCACAGT 300  
 OY 301 GTCCCTTATGTAAGATGATATATAACCAAAAGAGCCTACAGAAAGTACAGATTTAGTC 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 301 GTCCCTTATGTAAGATGATATATAACCAAAAGAGCCTACAGAAAGTACAGATTTAGTC 360  
 OY 361 AACTTGTGTAAGAGCTATTGAAATATCTTGTGCTTTTCACTTGAAGAGTTGGAGT 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 361 AACTTGTGTAAGAGCTATTGAAATATCTTGTGCTTTTCACTTGAAGAGTTGGAGT 420  
 OY 421 ATGCAAAACAGCTATATTTTTCAGAAAAGAAATTAACCTCTGTAACATCTTAAAGATG 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 421 ATGCAAAACAGCTATATTTTTCAGAAAAGAAATTAACCTCTGTAACATCTTAAAGATG 480  
 OY 481 AAGTTCCTATCATTCAAGATGAGGCTACAGAAACCGTCCAAAGAGCTCTACAGAGTG 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 481 AAGTTCCTATCATTCAAGATGAGGCTACAGAAACCGTCCAAAGAGCTCTACAGAGTG 540  
 OY 541 AACCCGAAATTCCTCTCTCCAGAGAAACAGTCTCAAGTGTCAACTCTTAACCTTGGA 600  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 541 AACCCGAAATTCCTCTCTCCAGAGAAACAGTCTCAAGTGTCAACTCTTAACCTTGGA 600  
 OY 601 CTGTGAGAACTCTGAGAGCAAGAGCGGATATACACTCAAAAGAGCTGTCTACATTG 660  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 601 CTGTGAGAACTCTGAGAGCAAGAGCGGATATACACTCAAAAGAGCTGTCTACATTG 660  
 OY 661 AATGGAGATCTGATCTTCTGAGAGATACCGTTAATAAGCAACTTATTCAGAGTGGAG 720  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 661 AATGGAGATCTGATCTTCTGAGAGATACCGTTAATAAGCAACTTATTCAGAGTGGAG 720

OY	721	ATCAGAAATGGTTACAAATCACCCCTCAAGACACAGGATGAAATCAGTTGGATTCTG	780
Db	721	ATCAAGAAATGGTTACAAATCACCCCTCAAGGACACAGGATGAAATCAGTTGGATTCTG	780
OY	781	CAAAAAGAGCTCTGTGTAATTTCTGAGACGAGTGTAACAAAATCTGAACTATCATC	840
Db	781	CAAAAAGAGCTCTGTGTAATTTCTGAGACGAGTGTAACAAAATCTGAACTATCATC	840
OY	841	CCAGTAATATGATTTTGAACACACACTGAGAAAGCGTGCAGCTGAGAGGATCCAGAAAGT	900
Db	841	CCAGTAATATGATTTTGAACACACACTGAGAAAGCGTGCAGCTGAGAGGATCCAGAAAGT	900
OY	901	ATCAGGGTAGTCTCTTTCAAACTTGACATGTGAGACCATGTGAGCAAAATACTCATGCCA	960
Db	901	ATCAGGGTAGTCTCTTTCAAACTTGACATGTGAGACCATGTGAGCAAAATACTCATGCCA	960
OY	961	GCTCATTTACAGCATAGAAACAGCAGCTTTATTTACTACATAAGACAGAAATGAATGTAGANA	1020
Db	961	GCTCATTTACAGCATAGAAACAGCAGCTTTATTTACTACATAAGACAGAAATGAATGTAGANA	1020
OY	1021	AGGCTGAATTTCTGTATTAAGCAACACAGCTTGGCTTAGCAAGAGCCACATTAACAGAT	1080
Db	1021	AGGCTGAATTTCTGTATTAAGCAACACAGCTTGGCTTAGCAAGAGCCACATTAACAGAT	1080
OY	1081	GGGCTGGAAGTAGGAACATGATATGATAGGCGGACTCCGACAGCAAAAAAAGGTAG	1140
Db	1081	GGGCTGGAAGTAGGAACATGATATGATAGGCGGACTCCGACAGCAAAAAAAGGTAG	1140
OY	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAATTAAGCAGAAACTGCCATGCT	1200
Db	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAATTAAGCAGAAACTGCCATGCT	1200
OY	1201	CAGAAATTCCTAGACATCTGAAAGTGTTCCTTGATTAACACTAAATAGCAGCATTTGAGA	1260
Db	1201	CAGAAATTCCTAGACATCTGAAAGTGTTCCTTGATTAACACTAAATAGCAGCATTTGAGA	1260
OY	1261	AAGTAAATGAGTGGTTTTCCAGAAATGATGAACCTTAGGTTCGTATGACTACATGATG	1320
Db	1261	AAGTAAATGAGTGGTTTTCCAGAAATGATGAACCTTAGGTTCGTATGACTACATGATG	1320
OY	1321	GGGAGTCTGAATCAATGCCCAAGTAGCTGATATTGGAGCTTAAATGAGTAGATG	1380
Db	1321	GGGAGTCTGAATCAATGCCCAAGTAGCTGATATTGGAGCTTAAATGAGTAGATG	1380
OY	1381	AATATTCCTGATCTTCAGAGAAATAGACTTACCTGGCAGAGTATCCTATGAGGCTTTAA	1440
Db	1381	AATATTCCTGATCTTCAGAGAAATAGACTTACCTGGCAGAGTATCCTATGAGGCTTTAA	1440
OY	1441	TATGTAAAGTGAAGAGTCTCACTCCAATCAGTAGAGAGTAAATTGTAAGACAAAAATAT	1500
Db	1441	TATGTAAAGTGAAGAGTCTCACTCCAATCAGTAGAGAGTAAATTGTAAGACAAAAATAT	1500
OY	1501	TTGGGAAAACCTATGCGAAGAAAGGCAACCTCCCACTTAAGCATGTAATCTGAAATC	1560
Db	1501	TTGGGAAAACCTATGCGAAGAAAGGCAACCTCCCACTTAAGCATGTAATCTGAAATC	1560
OY	1561	TAAATTATGAGAGCACTTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCACAAATA	1620
Db	1561	TAAATTATGAGAGCACTTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCACAAATA	1620
OY	1621	AATTAAACGTAAGAAAGAGACCTACATCAGAGCCTTCACTCGAGGATTTTATCAAGAAAG	1680
Db	1621	AATTAAACGTAAGAAAGAGACCTACATCAGAGCCTTCACTCGAGGATTTTATCAAGAAAG	1680
OY	1681	CAGATTTTGGCACTTCAAAAAGACTCCTGAAATGATTAATCAGAGAACTAACCAAGCGAGC	1740
Db	1681	CAGATTTTGGCACTTCAAAAAGACTCCTGAAATGATTAATCAGAGAACTAACCAAGCGAGC	1740
OY	1741	AGAAATGGTCAAGTATGAAATTTACTAATAGTGGTATGAGATTAACCAAAAGGATGTT	1800
Db	1741	AGAAATGGTCAAGTATGAAATTTACTAATAGTGGTATGAGATTAACCAAAAGGATGTT	1800

OY		1801	CTATTTCAGAAATGAGAATAATCTTAACCCCAATAGATCTCTGGAAAAAAGATCTCGTTTCA	1860
Db		1801	CTATTTCAGAAATGAGAATAATCTTAACCCCAATAGATCTCTGGAAAAAAGATCTCGTTTCA	1860
OY		1861	AAACGAAGCTGAAACCTTTAAAGCAGCAGCTATTAAGCAATATGTGAAGTGCATTTAATATATCC	1920
Db		1861	AAACGAAGCTGAAACCTTTAAAGCAGCAGCTATTAAGCAATATGTGAAGTGCATTTAATATATCC	1920
OY		1921	ACAATTCAAAAGCACCTTAAAAAGAAATAGGCTGAGAGGAGAACTCTTCCAAGCATATTC	1980
Db		1921	ACAATTCAAAAGCACCTTAAAAAGAAATAGGCTGAGAGGAGAACTCTTCCAAGCATATTC	1980
OY		1981	ATGCCCTTGAACTACTAGTACTAGTAAATCTTAAGCCCACCTTAATTTGTAATGCAATTCGAAA	2040
Db		1981	ATGCCCTTGAACTACTAGTACTAGTAAATCTTAAGCCCACCTTAATTTGTAATGCAATTCGAAA	2040
OY		2041	TTCGATGAGTGTCTTCGAGTAGAAGGATTAAGAAAAAAAGTACACCAAATGTCAGTCA	2100
Db		2041	TTCGATGAGTGTCTTCGAGTAGAAGGATTAAGAAAAAAAGTACACCAAATGTCAGTCA	2100
OY		2101	GGCACAGCAGAAACCTTACACTCATATGAGGTAAAGAACCTGCACACTGAGACCAGAAAGA	2160
Db		2101	GGCACAGCAGAAACCTTACACTCATATGAGGTAAAGAACCTGCACACTGAGACCAGAAAGA	2160
OY		2161	GTAAACAAGCCAAATGAAACAGACAGAAATGAAGACATGACATGATATCTTCCAGAGCTGA	2220
Db		2161	GTAAACAAGCCAAATGAAACAGACAGAAATGAAGACATGACATGATATCTTCCAGAGCTGA	2220
OY		2221	AGTTAACAAATGACACTGCTGTTTTCTTAAGTGTTCAAATACCAATGTAACCTTAAAGAT	2280
Db		2221	AGTTAACAAATGACACTGCTGTTTTCTTAAGTGTTCAAATACCAATGTAACCTTAAAGAT	2280
OY		2281	TTGTCAATCTAGCCCTTCAGAGAGAGAAAGAAAGAACTGAAACAGTTAAAGTGT	2340
Db		2281	TTGTCAATCTAGCCCTTCAGAGAGAGAAAGAAAGAACTGAAACAGTTAAAGTGT	2340
OY		2341	CTAATTAATGCTGAAGAGCCCCCAAAGATCTCATGTTAAGTGAGAAAGGGTTTTGCCAACTG	2400
Db		2341	CTAATTAATGCTGAAGAGCCCCCAAAGATCTCATGTTAAGTGAGAAAGGGTTTTGCCAACTG	2400
OY		2401	AAAGATCTGTAGAGAGTAGACAGTATTTTCATTGTGTAACCTGGTACTGATTTATGGCACTCAGG	2460
Db		2401	AAAGATCTGTAGAGAGTAGACAGTATTTTCATTGTGTAACCTGGTACTGATTTATGGCACTCAGG	2460
OY		2461	AAAGATCTGCTTACTGGAAGTTTGAACCTTGAAGGAGGCAAAAAACGAACCAATTAAT	2520
Db		2461	AAAGATCTGCTTACTGGAAGTTTGAACCTTGAAGGAGGCAAAAAACGAACCAATTAAT	2520
OY		2521	GTCGTAGTCAAGTGTGCACAGCTTTTAAAAACCCCAAGGGGCACTAATTCATGAGTTGTTCCAAG	2580
Db		2521	GTCGTAGTCAAGTGTGCACAGCTTTTAAAAACCCCAAGGGGCACTAATTCATGAGTTGTTCCAAG	2580
OY		2581	ATAATAGAAATGACACACAGAGGCTTTTAAGTATCCATTGGGACATGAAGTTAACCACAGTC	2640
Db		2581	ATAATAGAAATGACACACAGAGGCTTTTAAGTATCCATTGGGACATGAAGTTAACCACAGTC	2640
OY		2641	GGAAGAACAGCATGGAATGGAAGAAAGTGAAGTGAATGCACATATTTGGAGAAATCAT	2700
Db		2641	GGAAGAACAGCATGGAATGGAAGAAAGTGAAGTGAATGCACATATTTGGAGAAATCAT	2700
OY		2701	TCAAGGTTTCAAGAGCCGACAGTCAATTTGCTCTGTTTTCAATCCAGGAATGACAGAAAGG	2760
Db		2701	TCAAGGTTTCAAGAGCCGACAGTCAATTTGCTCTGTTTTCAATCCAGGAATGACAGAAAGG	2760
OY		2761	AATGTGCAACATTTCTGTGCCACACTCTGGTCTTTAAAGAAACAAAGTCCAAAAGTCACTT	2820
Db		2761	AATGTGCAACATTTCTGTGCCACACTCTGGTCTTTAAAGAAACAAAGTCCAAAAGTCACTT	2820
OY		2821	TTGAATGGAACAAGAGAAAGAAATATCAGGAAAGAAATGAGTCTAATATCAAGCTGTAC	2880
Db		2821	TTGAATGGAACAAGAGAAAGAAATATCAGGAAAGAAATGAGTCTAATATCAAGCTGTAC	2880
OY		2881	AGACAGTTTAATATCTGACAGGCTTCTCTGTGTTGGTGCAGAAAGATATACCCAGTTGATA	2940



```

OY 5101 AATTATGCTGCTGACAGTTTGGCAGAAAACACACATCATTAACTAATCTAATTA 5160
    |||||||
DB 5101 AATTATGCTGCTGACAGTTTGGCAGAAAACACACATCATTAACTAATCTAATTA 5160
OY 5161 CTGAAGAGACTACATGTTGTTATGAAAACAGATGCTGATTTGTGTGAACGAGAC 5220
    |||||||
DB 5161 CTGAAGAGACTACATGTTGTTATGAAAACAGATGCTGATTTGTGTGAACGAGAC 5220
OY 5221 TGAATATTTTCTAGGAATTTGGGAGGAAATGGGTAGTTAGCTATTTCTGGTGACC 5280
    |||||||
DB 5221 TGAATATTTTCTAGGAATTTGGGAGGAAATGGGTAGTTAGCTATTTCTGGTGACC 5280
OY 5281 AGCTATTAAAGAAAAGAAAATGCTGAATGACATGATTTTGAAGTCAGAGAGATGTGG 5340
    |||||||
DB 5281 AGCTATTAAAGAAAAGAAAATGCTGAATGACATGATTTTGAAGTCAGAGAGATGTGG 5340
OY 5341 TCAATGGAAGAAACACCAAGGTCCAAAAGCAGAGAAATCCAGAGACAGAAAAGATCT 5400
    |||||||
DB 5341 TCAATGGAAGAAACACCAAGGTCCAAAAGCAGAGAAATCCAGAGACAGAAAAGATCT 5400
OY 5401 TCAGGGGGCTGAATCTGTTGCTATGGGCTTGCACCAACATGCCCCACAGATCAACTGG 5460
    |||||||
DB 5401 TCAGGGGGCTGAATCTGTTGCTATGGGCTTGCACCAACATGCCCCACAGATCAACTGG 5460
OY 5461 AATGATGATGACAGCTGTGTGCTGCTGTGTGTAAGAGCTTTCATCATTCACCCCTTG 5520
    |||||||
DB 5461 AATGATGATGACAGCTGTGTGCTGCTGTGTGTAAGAGCTTTCATCATTCACCCCTTG 5520
OY 5461 AATGATGATGACAGCTGTGTGCTGCTGTGTGTAAGAGCTTTCATCATTCACCCCTTG 5520
    |||||||
DB 5461 AATGATGATGACAGCTGTGTGCTGCTGTGTGTAAGAGCTTTCATCATTCACCCCTTG 5520
OY 5521 GCACAGGTGTCCACCAATTTGTTGTGTGTCAGACAGATGCTTGACAGAGCAATGCT 5580
    |||||||
DB 5521 GCACAGGTGTCCACCAATTTGTTGTGTGTCAGACAGATGCTTGACAGAGCAATGCT 5580
OY 5581 TCCATGCAATGGGAGATGTGTGAGGACCTGTGTGTACCCGAAAGTGGGTGTGGACA 5640
    |||||||
DB 5581 TCCATGCAATGGGAGATGTGTGAGGACCTGTGTGTACCCGAAAGTGGGTGTGGACA 5640
OY 5641 GTGTAGCACTACAGTGCAGAGAGCTGAGACCTACTGATACCCAGATCCCCCACA 5700
    |||||||
DB 5641 GTGTAGCACTACAGTGCAGAGAGCTGAGACCTACTGATACCCAGATCCCCCACA 5700
OY 5701 GCCACTACTGA 5711
    |||||||
DB 5701 GCCACTACTGA 5711

RESULT 9
ARO04673 5914 bp DNA linear PAT 04-DEC-1998
LOCUS Sequence 1 from patent US 5747282.
DEFINITION ARO04673
ACCESSION ARO04673
VERSION AR004673.1 GI:3965552
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 5914)
  Skolnick M.H., Goldgar D.E., Miki Y., Swenson J., Kanb A.,
  Harshman K.D., Shattuck-Eidens D.M., Tavtiglian S.V., Wiseman R.W.
  and Futreal P. Andrew.
  170-Linked breast and ovarian cancer susceptibility gene
  Patent: US 5747282-A 1 05-MAY-1998;
  Location/Qualifiers
    source 1..5914
    ORGANISM="Unknown"
BASE COUNT 2006 a 1156 c 1316 g 1436 t
ORIGIN
Query Match 99.9%; Score 5707.8; DB 6; Length 5914;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AGCTGCTGAGACTTCTGAGACCCCGCAGAGGCTGTGGGCTTCTCAGATAACTGGGCC 60

```

```

DB 1 AGCTGCTGAGACTTCTGAGACCCCGCAGAGGCTGTGGGCTTCTCAGATAACTGGGCC 60
    |||||||
OY 61 CCTGGCGTCAGAGAGCCTTACCCCTCTGCTGTGGTAAAGTTCAATGGACGAAGAANA 120
    |||||||
DB 61 CCTGGCGTCAGAGAGCCTTACCCCTCTGCTGTGGTAAAGTTCAATGGACGAAGAANA 120
OY 121 TCGATTTATCTCTCTGCGGCTTGAAGAGATCAAAATGTCATTAATGCTATGCGAANA 180
    |||||||
DB 121 TCGATTTATCTCTCTGCGGCTTGAAGAGATCAAAATGTCATTAATGCTATGCGAANA 180
OY 181 TCTTGAAGTCCCATCTCTCTGAGATTGATCAAGAACTCTCTCCAAAGTGTGACC 240
    |||||||
DB 181 TCTTGAAGTCCCATCTCTCTGAGATTGATCAAGAACTCTCTCCAAAGTGTGACC 240
OY 240 TCTTGAAGTCCCATCTCTCTGAGATTGATCAAGAACTCTCTCCAAAGTGTGACC 240
    |||||||
DB 240 TCTTGAAGTCCCATCTCTCTGAGATTGATCAAGAACTCTCTCCAAAGTGTGACC 240
OY 241 ACATATTTTGCAGAAATTTTGCATGCTGAAACTTCTACACAGAGAAAGGCTTACAGT 300
    |||||||
DB 241 ACATATTTTGCAGAAATTTTGCATGCTGAAACTTCTACACAGAGAAAGGCTTACAGT 300
OY 301 GTCCTTTATGTAAGATGATATACCAAAAGAGCCTTACAGAAAGTACAGATTTAGTC 360
    |||||||
DB 301 GTCCTTTATGTAAGATGATATACCAAAAGAGCCTTACAGAAAGTACAGATTTAGTC 360
OY 361 AACTGTTGAAGAGCTATTTGAAATCATTTGCTTTTCAGCTTGACACAGGTTTGAAGT 420
    |||||||
DB 361 AACTGTTGAAGAGCTATTTGAAATCATTTGCTTTTCAGCTTGACACAGGTTTGAAGT 420
OY 421 ATGCAAAACGCTATATTTTGCAGAAAGGAAATTAATCTCTCTAATCACTTAAAGATG 480
    |||||||
DB 421 ATGCAAAACGCTATATTTTGCAGAAAGGAAATTAATCTCTCTAATCACTTAAAGATG 480
OY 481 AAGTTTATCATCTCAAAAGTATGGGCTACAGAAACCGTCCAAAGAGCTTACAGAGTG 540
    |||||||
DB 481 AAGTTTATCATCTCAAAAGTATGGGCTACAGAAACCGTCCAAAGAGCTTACAGAGTG 540
OY 541 AACCAGAAATCTCTCTGAGAGAAACAGTCTAGTGTCCAACTCTTAACTTTGANA 600
    |||||||
DB 541 AACCAGAAATCTCTCTGAGAGAAACAGTCTAGTGTCCAACTCTTAACTTTGANA 600
OY 601 CTGTAGAGACTGTGAGGACAAAGAGCGGATTCAACTCTCAAAAGCGTGTCTACATTTG 660
    |||||||
DB 601 CTGTAGAGACTGTGAGGACAAAGAGCGGATTCAACTCTCAAAAGCGTGTCTACATTTG 660
OY 661 AATGGAGATGATCTCTCTGAGAGATACCGTTAATTAAGGCAACTTATTTGAGTGGAG 720
    |||||||
DB 661 AATGGAGATGATCTCTCTGAGAGATACCGTTAATTAAGGCAACTTATTTGAGTGGAG 720
OY 721 ATCAAGAAATTTTACAAATCACCCCTCAAGGAACGAGGATGAATTCAGTTTGAATCTG 780
    |||||||
DB 721 ATCAAGAAATTTTACAAATCACCCCTCAAGGAACGAGGATGAATTCAGTTTGAATCTG 780
OY 781 CAAAAAAGCGCTGTGTGAAATTTCTGAGAGCGATGAACAAATCTCAACATCTCAAC 840
    |||||||
DB 781 CAAAAAAGCGCTGTGTGAAATTTCTGAGAGCGATGAACAAATCTCAACATCTCAAC 840
OY 841 CCAGTAATTAATGATTTGAACACCACTGAGAGCGGAGCGTGAAGGATCCAGAAAAGT 900
    |||||||
DB 841 CCAGTAATTAATGATTTGAACACCACTGAGAGCGGAGCGTGAAGGATCCAGAAAAGT 900
OY 901 ATCAGAGTAGTCTGTTCAAACTGCAATGAGGACCAATGAGGCAATTAATCTATGCGCA 960
    |||||||
DB 901 ATCAGAGTAGTCTGTTCAAACTGCAATGAGGACCAATGAGGCAATTAATCTATGCGCA 960
OY 961 GCTCATTTACAGATGAGAGACAGCACTTTATTAAGTAAAGAGCAATTAATGATGAGAAA 1020
    |||||||
DB 961 GCTCATTTACAGATGAGAGACAGCACTTTATTAAGTAAAGAGCAATTAATGATGAGAAA 1020
OY 1021 AGGCTGAATTTCTGTAATAAAGCAAGACCGGCTTGAAGAGGACCAATTAATCAAGAT 1080
    |||||||
DB 1021 AGGCTGAATTTCTGTAATAAAGCAAGACCGGCTTGAAGAGGACCAATTAATCAAGAT 1080
OY 1081 GGGCTGAAGATTAAGAAACATGTAATGATAGGCGGACTCCAGACAGAAAAAAGCTAG 1140
    |||||||
DB 1081 GGGCTGAAGATTAAGAAACATGTAATGATAGGCGGACTCCAGACAGAAAAAAGCTAG 1140

```

Db 1081 GGGCTGGAAGTAAAGAAACATCTAATAGTGGCGAGCTCCGACGACAGAAAAAAGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAAGTAATAGCAGAACTGCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAAGTAATAGCAGAACTGCCATGCT 1200  
QY 1201 CAGAGAACTCTAGAGATCTGAGAGATGTTCTTGATATACACTAATAATAGCAGATTCAGA 1260  
Db 1201 CAGAGAACTCTAGAGATCTGAGAGATGTTCTTGATATACACTAATAATAGCAGATTCAGA 1260  
QY 1261 AAGTAATGAGTGGTTCCTCCAGAGTGAATGTAAGTGGTTCGATGATCAGCATGATG 1320  
Db 1261 AAGTAATGAGTGGTTCCTCCAGAGTGAATGTAAGTGGTTCGATGATCAGCATGATG 1320  
QY 1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATTGGAGCTTCTAATAGTAGATG 1380  
Db 1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATTGGAGCTTCTAATAGTAGATG 1380  
QY 1381 AATATCTGCTTCTTCAGAGAAATAGACTTACTGGCCAGTGAATCCTCATGAGCTTTAA 1440  
Db 1381 AATATCTGCTTCTTCAGAGAAATAGACTTACTGGCCAGTGAATCCTCATGAGCTTTAA 1440  
QY 1441 TATGTAAAGTGAAGAGGTTCACTCCAAATCAGTAGAGATTAATTAAGACAAATAT 1500  
Db 1441 TATGTAAAGTGAAGAGGTTCACTCCAAATCAGTAGAGATTAATTAAGACAAATAT 1500  
QY 1501 TTGGGAAACCTTATCGGAAAGAGGCAAGCCCTCCCACTTAAGCCATTAAGTGAATC 1560  
Db 1501 TTGGGAAACCTTATCGGAAAGAGGCAAGCCCTCCCACTTAAGCCATTAAGTGAATC 1560  
QY 1561 TAAATATAGAGCACTTTTGTACTGAGCCACAGATTAATCAAGAGCTCCCTCACAATA 1620  
Db 1561 TAAATATAGAGCACTTTTGTACTGAGCCACAGATTAATCAAGAGCTCCCTCACAATA 1620  
QY 1621 AATTAAGCGTAAAGAGAGACCTCATCAGGCTTCATCCGAGATTTTTCAGAAAG 1680  
Db 1621 AATTAAGCGTAAAGAGAGACCTCATCAGGCTTCATCCGAGATTTTTCAGAAAG 1680  
QY 1681 CAGATTTGGCAGTTCAAAAGACTCCTGAATATGATTAATCAGGAACTAACCAAGGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAGACTCCTGAATATGATTAATCAGGAACTAACCAAGGAGC 1740  
QY 1741 AGAATGCTCAAGTGAATATTTACTAATAGTGGTCAATGAGATTAATCAAAAGGATG 1800  
Db 1741 AGAATGCTCAAGTGAATATTTACTAATAGTGGTCAATGAGATTAATCAAAAGGATG 1800  
QY 1801 CTATTCAGATGAGAAATCTCTAACCCATAGAAATCAGTGAATTAATATATC 1860  
Db 1801 CTATTCAGATGAGAAATCTCTAACCCATAGAAATCAGTGAATTAATATATC 1860  
QY 1861 AAACGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGGAATCGAATTAATATC 1920  
Db 1861 AAACGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGGAATCGAATTAATATC 1920  
QY 1921 ACAATTTCAAAAGCACTTAAGAAATAGGCTGAGGAGAGAGTCTTTCTACAGGCAATTC 1980  
Db 1921 ACAATTTCAAAAGCACTTAAGAAATAGGCTGAGGAGAGAGTCTTTCTACAGGCAATTC 1980  
QY 1981 ATGCGCTTGAATAGTAGTAGTAGTAAGATCTAAGCCACACCTAATTTGACTGAATTTGAAA 2040  
Db 1981 ATGCGCTTGAATAGTAGTAGTAGTAAGATCTAAGCCACACCTAATTTGACTGAATTTGAAA 2040  
QY 2041 TTGATAGTGTCTTACAGTAGTAAGAGATTAAGAAAAAAGTACAAACCAATGCGACATCA 2100  
Db 2041 TTGATAGTGTCTTACAGTAGTAAGAGATTAAGAAAAAAGTACAAACCAATGCGACATCA 2100  
QY 2101 GGCACAGCAGAAACCTTAACACTCATGGAAGTAAAGAACTGCAACTGGAGCCAAAGAGA 2160  
Db 2101 GGCACAGCAGAAACCTTAACACTCATGGAAGTAAAGAACTGCAACTGGAGCCAAAGAGA 2160  
QY 2161 GTACAGTGAAGCAATTAAGAGAGAGAGTAAAGAGATGACATGACTTTCCAGAGCTGA 2220  
Db 2161 GTACAGTGAAGCAATTAAGAGAGAGAGTAAAGAGATGACATGACTTTCCAGAGCTGA 2220

QY 2221 AGTTAACAAATGCACCTGGTCTTTTACTAAGTGTCAAAATACAGTAGTAATTAAGAAAT 2280  
Db 2221 AGTTAACAAATGCACCTGGTCTTTTACTAAGTGTCAAAATACAGTAGTAATTAAGAAAT 2280  
QY 2281 TTGTCAATCTGACCTTCCAGAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGTGT 2340  
Db 2281 TTGTCAATCTGACCTTCCAGAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGTGT 2340  
QY 2341 CTAAATAGTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGTTTCCAAACTG 2400  
Db 2341 CTAAATAGTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGTTTCCAAACTG 2400  
QY 2401 AAAGATCTGAGAGATAGCAGTATTTCAATGTTGACCTGATGATGATGATGATGATG 2460  
Db 2401 AAAGATCTGAGAGATAGCAGTATTTCAATGTTGACCTGATGATGATGATGATGATG 2460  
QY 2461 AAAGATCTGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 2520  
Db 2461 AAAGATCTGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATG 2520  
QY 2521 GTGTGAGTCACTGTCAGCAGTATTTGAAACCCCAAGGACTAATTCATGTTTCCAAAG 2580  
Db 2521 GTGTGAGTCACTGTCAGCAGTATTTGAAACCCCAAGGACTAATTCATGTTTCCAAAG 2580  
QY 2581 ATATATAGAAATGACAGAGAGAGCTTTAAGTATCCATGAGACATGAAGTTAACACAGTC 2640  
Db 2581 ATATATAGAAATGACAGAGAGAGCTTTAAGTATCCATGAGACATGAAGTTAACACAGTC 2640  
QY 2641 GGGAAACAGCATTAAGAAATGAGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2700  
Db 2641 GGGAAACAGCATTAAGAAATGAGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2700  
QY 2701 TCAAGGTTTCAAGCGCAGTATGCTGCTGTTTCAATCCAGAGAAATGCAAGAGAG 2760  
Db 2701 TCAAGGTTTCAAGCGCAGTATGCTGCTGTTTCAATCCAGAGAAATGCAAGAGAG 2760  
QY 2761 AATGTGCAACATTCCTGCCCCACTGCTGCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
Db 2761 AATGTGCAACATTCCTGCCCCACTGCTGCTTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
QY 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGGAAAGTGAATGATCAATCAACCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGGAAAGTGAATGATCAATCAACCTGTAC 2880  
QY 2881 AGACAGTAAATATACAGAGGCTTCCGTGGTGTGTCAGAAAGTAAAGCCAGTTGATA 2940  
Db 2881 AGACAGTAAATATACAGAGGCTTCCGTGGTGTGTCAGAAAGTAAAGCCAGTTGATA 2940  
QY 2941 ATGCAATATGATACAGAGGCTCTAGGTTTGTCTATCATCTCAGTTTCAGAGGCA 3000  
Db 2941 ATGCAATATGATACAGAGGCTCTAGGTTTGTCTATCATCTCAGTTTCAGAGGCA 3000  
QY 3001 ACAGAACTGACATCTTACCTCAAAATTAACATGAGTATTAACAAACCATATCTGATAC 3060  
Db 3001 ACAGAACTGACATCTTACCTCAAAATTAACATGAGTATTAACAAACCATATCTGATAC 3060  
QY 3061 CACCACTTTTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3120  
Db 3061 CACCACTTTTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3120  
QY 3121 AAAACCTTGAAGAAACATTCATGATGATGATGATGATGATGATGATGATGATGATG 3180  
Db 3121 AAAACCTTGAAGAAACATTCATGATGATGATGATGATGATGATGATGATGATGATG 3180  
QY 3181 GTACAGTGAAGCAATTAAGAGAGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAG 3240  
Db 3181 GTACAGTGAAGCAATTAAGAGAGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
QY 3241 CAAGCAATTAATTAAGAGAGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Db 3241 CAAGCAATTAATTAAGAGAGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300



QY	3301	TAGGTTCCAGTGGATGAAACATTCACAGCAGACTAGTAGGAAACAGAGGGCCCAAAATTGCA	3360
Db	3301	TAGGTTCCAGTGGATGAAACATTCACAGCAGACTAGTAGGAAACAGAGGGCCCAAAATTGCA	3360
QY	3361	ATGCTATGCTTAGATTTAGGGGTTTGGCAACCTGAGGCTATATAACCAAGATCTTCCGGAA	3420
Db	3361	ATGCTATGCTTAGATTTAGGGGTTTGGCAACCTGAGGCTATATAACCAAGATCTTCCGGAA	3420
QY	3421	GTAATTGTAAGCATCTCGAATAATAAAAAGCAGAATAATGAGAGAGTAGTTCCAGACTGTTTA	3480
Db	3421	GTAATTGTAAGCATCTCGAATAATAAAAAGCAGAATAATGAGAGAGTAGTTCCAGACTGTTTA	3480
QY	3481	ATACAGATTTCTCTCCATATCTGATTTCCAGATTACTTTAGAACGCTTATGGAGAGTACTG	3540
Db	3481	ATACAGATTTCTCTCCATATCTGATTTCCAGATTACTTTAGAACGCTTATGGAGAGTACTG	3540
QY	3541	ATGATCTCAGGTTTGTCTGAGACACCTGATACCTGTTAGATGATGTTGTAATAAAGG	3600
Db	3541	ATGATCTCAGGTTTGTCTGAGACACCTGATACCTGTTAGATGATGTTGTAATAAAGG	3600
QY	3601	AAGTACTAGTTTTGGCTGGAATAATGACATTTAAGSAAAGTTCTGCTGTTTATACAAAGCG	3660
Db	3601	AAGTACTAGTTTTGGCTGGAATAATGACATTTAAGSAAAGTTCTGCTGTTTATACAAAGCG	3660
QY	3661	TCCAGAAAGAGAGACTTAGCAGAGAGCTTAGCCCTTTCACCCATACATATTTGGGCTCAGG	3720
Db	3661	TCCAGAAAGAGAGACTTAGCAGAGAGCTTAGCCCTTTCACCCATACATATTTGGGCTCAGG	3720
QY	3721	GTTTCCGCAAGAGGGGCCAAGAAATTAGAGCTCTCAGAGAGAACTTATCTAGTAGAGATG	3780
Db	3721	GTTTCCGCAAGAGGGGCCAAGAAATTAGAGCTCTCAGAGAGAACTTATCTAGTAGAGATG	3780
QY	3781	AAGAGCTTCCCTGCTTCCAACTGTTATTGGTAAAGTAAACATATACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTTCCAACTGTTATTGGTAAAGTAAACATATACCTTCTCAGT	3840
QY	3841	CTACTAGGCAATAGACCCGTTGTCTACCGAGTGTCTGTCTAAGAACACAGAGGAGAAATTTAT	3900
Db	3841	CTACTAGGCAATAGACCCGTTGTCTACCGAGTGTCTGTCTAAGAACACAGAGGAGAAATTTAT	3900
QY	3901	TATCATTTGAAGAAATAGCTTAATAGCTGACGATACCGAGTAATATTGGCAAGGATCTC	3960
Db	3901	TATCATTTGAAGAAATAGCTTAATAGCTGACGATACCGAGTAATATTGGCAAGGATCTC	3960
QY	3961	AGGAACATCACTTAGTAGAGGAACAAAAATGTTCTAGCTTGTTTTCTTACACGTGCA	4020
Db	3961	AGGAACATCACTTAGTAGAGGAACAAAAATGTTCTAGCTTGTTTTCTTACACGTGCA	4020
QY	4021	GTTGATTTGGAAGACTGCTGCAATAATACAAACACCAGAGATCCTTCTTGATTTGGTTCTT	4080
Db	4021	GTTGATTTGGAAGACTGCTGCAATAATACAAACACCAGAGATCCTTCTTGATTTGGTTCTT	4080
QY	4081	CCAAACAAATGAGGCACTCTGAAAGCCAGGAGATTGGTCTGAGTGCACAGGAATTGG	4140
Db	4081	CCAAACAAATGAGGCACTCTGAAAGCCAGGAGATTGGTCTGAGTGCACAGGAATTGG	4140
QY	4141	TTTTCAGATGATGGAAGAGAGAGAGCGGCTTGGAAAGAAATATATCAAGAGAGCAAGCA	4200
Db	4141	TTTTCAGATGATGGAAGAGAGAGAGCGGCTTGGAAAGAAATATATCAAGAGAGCAAGCA	4200
QY	4201	TGATTTCAAACTTAGTGAAGAGCAGACTTGGGTGTGAGAGTGAACCAACGCTCTCTGAAG	4260
Db	4201	TGATTTCAAACTTAGTGAAGAGCAGACTTGGGTGTGAGAGTGAACCAACGCTCTCTGAAG	4260
QY	4261	ACTGCTCAGGGGCTATCTCTCAAGATGACATTTTAAACCTCAGCAGAGAGGATACCATGC	4320
Db	4261	ACTGCTCAGGGGCTATCTCTCAAGATGACATTTTAAACCTCAGCAGAGAGGATACCATGC	4320
QY	4321	AACATTAACCTGATTAACCTCAGAGAGAAATGGCTGACAGTAAAGAGTGTGTTAGAAACGC	4380
Db	4321	AACATTAACCTGATTAACCTCAGAGAGAAATGGCTGACAGTAAAGAGTGTGTTAGAAACGC	4380
QY	4381	ATGGAGCCAGGCTTCTTAACAGCTACCTTCCATCAATAGTGACTTCTTGCCCTTGAGG	4440
Db	4381	ATGGAGCCAGGCTTCTTAACAGCTACCTTCCATCAATAGTGACTTCTTGCCCTTGAGG	4440
QY	4441	ACCTGCGAAATCCAGAAACAGACATTCAGAAAGAACAGTATTTACCTTACACAGAAACGTA	4500
Db	4441	ACCTGCGAAATCCAGAAACAGACATTCAGAAAGAACAGTATTTACCTTACACAGAAACGTA	4500
QY	4501	GTCGAATACCTTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTTGAGTGTCTG	4560
Db	4501	GTCGAATACCTTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTTGAGTGTCTG	4560
QY	4561	CAGATAGTCTTACCAGTAAAAATTAAGAACACAGAGTGGAAAGGTATCTCCCTTTAAAT	4620
Db	4561	CAGATAGTCTTACCAGTAAAAATTAAGAACACAGAGTGGAAAGGTATCTCCCTTTAAAT	4620
QY	4621	GCCCATCATTAAGATGATAGGTGGTACATGACAGTGGCTGAGGAGTCTTCAGATTAAGA	4680
Db	4621	GCCCATCATTAAGATGATAGGTGGTACATGACAGTGGCTGAGGAGTCTTCAGATTAAGA	4680
QY	4681	ACTTACCANTCTCAAGAGAGAGCTTCATTAAGTGTGTTGATGTGAGAGAGCAACAGCTGAGAG	4740
Db	4681	ACTTACCANTCTCAAGAGAGAGCTTCATTAAGTGTGTTGATGTGAGAGAGCAACAGCTGAGAG	4740
QY	4741	AGTCTGGGCCACACGATTTGACGGAACATCTTACTTGCACAGGCAAGATCTAGAGGAA	4800
Db	4741	AGTCTGGGCCACACGATTTGACGGAACATCTTACTTGCACAGGCAAGATCTAGAGGAA	4800
QY	4801	CCCCCTTACCTGGAATCTGGAATCAGCTCTCTCTGATGATACCTGAAATCTGATCTCTG	4860
Db	4801	CCCCCTTACCTGGAATCTGGAATCAGCTCTCTCTGATGATACCTGAAATCTGATCTCTG	4860
QY	4861	AAGACAGAGCCCGAGAGTCAAGCTGCTGTTGGCAACATACCATCTTCAACCTGCTGATTA	4920
Db	4861	AAGACAGAGCCCGAGAGTCAAGCTGCTGTTGGCAACATACCATCTTCAACCTGCTGATTA	4920
QY	4921	AAAGTCCCAATTGAAGTTGCAGAAATCTGCCCCAGAGTCCAGCTGCTGCTCACTACTAG	4980
Db	4921	AAAGTCCCAATTGAAGTTGCAGAAATCTGCCCCAGAGTCCAGCTGCTGCTCACTACTAG	4980
QY	4981	ATTCGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAAGCCAGAAATTTGACAG	5040
Db	4981	ATTCGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAAGCCAGAAATTTGACAG	5040
QY	5041	CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGGTGTGCTGTGCTGACCCAGAG	5100
Db	5041	CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGGTGTGCTGTGCTGACCCAGAG	5100
QY	5101	AAATTATGCTGCTGTACAAAGTTGCCAGAAACACACATCACTTTAACTAATCTAATTA	5160
Db	5101	AAATTATGCTGCTGTGTACAAAGTTGCCAGAAACACACATCACTTTAACTAATCTAATTA	5160
QY	5161	CTGAAAGACTACTCATGTTGTATATGAAGAACAGATCCTGAGTGTGTGTGAACGAGAC	

Db 5461 AATGATGTAGACAGCTGTGTGCTCTGTGTGTGAAGAGACCTTTCATTCATTCACCCCTTG 5520  
 Oy 5521 GCACAGGTGTCCACCCCAATTTGTGTGTGACAGCAGATGCTGTGACAGAGACATGGCT 5580  
 Db 5521 GCACAGGTGTCCACCCCAATTTGTGTGTGACAGCAGATGCTGTGACAGAGACATGGCT 5580  
 Oy 5581 TCCATGCAATTTGGGACAGATGTGTGAGGACCTGTGTGTGACCCGAGAGTGGTGTGACA 5640  
 Db 5581 TCCATGCAATTTGGGACAGATGTGTGAGGACCTGTGTGTGACCCGAGAGTGGTGTGACA 5640  
 Oy 5641 GTGTAGACACTCTACCAATGCGCAGAGAGCTGTGACACCTTACTATACCCCAATGCCCA 5700  
 Db 5641 GTGTAGACACTCTACCAATGCGCAGAGAGCTGTGACACCTTACTATACCCCAATGCCCA 5700  
 Oy 5701 GCCACTACTGA 5711  
 Db 5701 GCCACTACTGA 5711

RESULT 10  
 LOCUS AR008159 5914 bp DNA Linear PAT 04-DEC-1998  
 DEFINITION Sequence 1 from patent US 5753441.  
 ACCESSION AR008159  
 VERSION AR008159.1 GI:3967268  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 5914)  
 AUTHORS Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
 Hershman,K.D., Shattuck-Bidens,D.M., Tavliglian,S.V., Wiseman,R.W.  
 and Futreal,P.Andrew.  
 TITLE 170-linked breast and ovarian cancer susceptibility gene  
 JOURNAL Patent: US 5753441-A 1 19-MAY-1998;  
 FEATURES  
 source 1..5914  
 location/Qualifiers  
 BASE COUNT 2006 a 1156 c 1316 g 1436 t  
 ORIGIN

Query Match 99.9%; Score 5707.8; DB 6; Length 5914;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 AGCTCGCTGAGACTTTCCTGACCCCGCAGCAGGCTGTGGGTTTCTCAGATACTGGGCC 60  
 Db 1 AGCTCGCTGAGACTTTCCTGACCCCGCAGCAGGCTGTGGGTTTCTCAGATACTGGGCC 60  
 Oy 61 CCTGCGCTCAGAGAGGCTTACCCCTGCTCTGCGTAAAGTTTCATTGGAAACAGAAAGAA 120  
 Db 61 CCTGCGCTCAGAGAGGCTTACCCCTGCTCTGCGTAAAGTTTCATTGGAAACAGAAAGAA 120  
 Oy 121 TGGATTATCTGCTCTTCGCGTTGAGAGAGTACAAAATGTCATTAAATGTCAGAAAA 180  
 Db 121 TGGATTATCTGCTCTTCGCGTTGAGAGAGTACAAAATGTCATTAAATGTCAGAAAA 180  
 Oy 121 TGGATTATCTGCTCTTCGCGTTGAGAGAGTACAAAATGTCATTAAATGTCAGAAAA 180  
 Db 121 TGGATTATCTGCTCTTCGCGTTGAGAGAGTACAAAATGTCATTAAATGTCAGAAAA 180  
 Oy 181 TCTTAAGTGTCCCATCTGTCTGAGTGTGATCAAGAACTGTCTCCACAAGGTGACC 240  
 Db 181 TCTTAAGTGTCCCATCTGTCTGAGTGTGATCAAGAACTGTCTCCACAAGGTGACC 240  
 Oy 181 TCTTAAGTGTCCCATCTGTCTGAGTGTGATCAAGAACTGTCTCCACAAGGTGACC 240  
 Db 181 TCTTAAGTGTCCCATCTGTCTGAGTGTGATCAAGAACTGTCTCCACAAGGTGACC 240  
 Oy 241 ACATATATTTGCAAAATTTTGCATGCTGAACTTCTCAACAGAGAAAGGCGCTTACAGT 300  
 Db 241 ACATATATTTGCAAAATTTTGCATGCTGAACTTCTCAACAGAGAAAGGCGCTTACAGT 300  
 Oy 301 GTTCCTTATGTAGATGATATATACCAAAAGAGAGCTTACAAAGATGATGATGTC 360  
 Db 301 GTTCCTTATGTAGATGATATATACCAAAAGAGAGCTTACAAAGATGATGATGTC 360  
 Oy 301 GTTCCTTATGTAGATGATATATACCAAAAGAGAGCTTACAAAGATGATGATGTC 360  
 Db 301 GTTCCTTATGTAGATGATATATACCAAAAGAGAGCTTACAAAGATGATGATGTC 360  
 Oy 361 AACTGTGTGAAGAGATTTGATGAATCATTTGTCTTTCAGTTGACACAGGTTTGAGT 420  
 Db 361 AACTGTGTGAAGAGATTTGATGAATCATTTGTCTTTCAGTTGACACAGGTTTGAGT 420  
 Oy 420 AACTGTGTGAAGAGATTTGATGAATCATTTGTCTTTCAGTTGACACAGGTTTGAGT 420  
 Db 420 AACTGTGTGAAGAGATTTGATGAATCATTTGTCTTTCAGTTGACACAGGTTTGAGT 420

Oy 421 ATGCAACAGCTATATTTTGCAAAAAAGAAATAAATCTCTCTGTAACATCTAAAGANG 480  
 Db 421 ATGCAACAGCTATATTTTGCAAAAAAGAAATAAATCTCTCTGTAACATCTAAAGANG 480  
 Oy 481 AAGTTTCTATCATCCAAAGATATGGGCTTACAGAAACGCTGCCAAAAGACTTCTACAGATG 540  
 Db 481 AAGTTTCTATCATCCAAAGATATGGGCTTACAGAAACGCTGCCAAAAGACTTCTACAGATG 540  
 Oy 541 AACCAGAAATCCCTCTCTGAGAGAAACAGTCTCACTGTCCAACTCTCAACCTTGAA 600  
 Db 541 AACCAGAAATCCCTCTCTGAGAGAAACAGTCTCACTGTCCAACTCTCAACCTTGAA 600  
 Oy 601 CTGTAGAACTCTGAGAGAAAGCAGGATATACACCTCAAAAGAGCTGTCTACATTTG 660  
 Db 601 CTGTAGAACTCTGAGAGAAAGCAGGATATACACCTCAAAAGAGCTGTCTACATTTG 660  
 Oy 661 AATTGGATCTGATCTTCTTGAAGATACCGTTAATAAGCACTTATTGACGTGTGGAG 720  
 Db 661 AATTGGATCTGATCTTCTTGAAGATACCGTTAATAAGCACTTATTGACGTGTGGAG 720  
 Oy 721 ATCAAGAAATGTATCAAAATCACCCCTCAAGGACCAAGGATGAATCAGTTTGATTTG 780  
 Db 721 ATCAAGAAATGTATCAAAATCACCCCTCAAGGACCAAGGATGAATCAGTTTGATTTG 780  
 Oy 781 CAAAAAAGGCTGCTTGTGAATTTTCTGAGAGGAGATGAACAAATACTGAACATCATCAAC 840  
 Db 781 CAAAAAAGGCTGCTTGTGAATTTTCTGAGAGGAGATGAACAAATACTGAACATCATCAAC 840  
 Oy 841 CCACTAATTAATGATTTTAACACACACAGAGGCTGAGCTGAGAGGATCCAGAAAGT 900  
 Db 841 CCACTAATTAATGATTTTAACACACACAGAGGCTGAGCTGAGAGGATCCAGAAAGT 900  
 Oy 901 ATCAGGCTAGTCTGTTCAAACTGTGATGTGAGGACATGTGACCAAAATACTCATGCCA 960  
 Db 901 ATCAGGCTAGTCTGTTCAAACTGTGATGTGAGGACATGTGACCAAAATACTCATGCCA 960  
 Oy 961 GCTCATTACAGCATGAGACAGCAGTTTATCTCACTAAGACAGATGAATGTAGAA 1020  
 Db 961 GCTCATTACAGCATGAGACAGCAGTTTATCTCACTAAGACAGATGAATGTAGAA 1020  
 Oy 1021 AGCTGAAATCTGTAATTAAGCAACAGCCTGGCTTAGCAGAGAGCCCAATTAACAGAT 1080  
 Db 1021 AGCTGAAATCTGTAATTAAGCAACAGCCTGGCTTAGCAGAGAGCCCAATTAACAGAT 1080  
 Oy 1081 GGGCTGGAAGTGAAGAAACATGTATGATGAGGAGATCCACACAGAAAAAAGGTAG 1140  
 Db 1081 GGGCTGGAAGTGAAGAAACATGTATGATGAGGAGATCCACACAGAAAAAAGGTAG 1140  
 Oy 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGGAATAGCAGAAACTGCCATGCT 1200  
 Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGGAATAGCAGAAACTGCCATGCT 1200  
 Oy 1201 CAGAGAACTCTAGAGATGATGAAAGTGTCTTGGATTAACACTAAATACACATTCACA 1260  
 Db 1201 CAGAGAACTCTAGAGATGATGAAAGTGTCTTGGATTAACACTAAATACACATTCACA 1260  
 Oy 1261 AAGTTAATGATGCTTTCCAGAAAGTGAAGTATGATGATGATGATGATGATGATGATG 1320  
 Db 1261 AAGTTAATGATGCTTTCCAGAAAGTGAAGTATGATGATGATGATGATGATGATGATG 1320  
 Oy 1321 GGGAGTCTGAATCAAAATGCCAAAGTACCTGATGATGATGATGATGATGATGATGATG 1380  
 Db 1321 GGGAGTCTGAATCAAAATGCCAAAGTACCTGATGATGATGATGATGATGATGATGATG 1380  
 Oy 1381 AATATCTGCTGCTTCTGAGAGAAATAGACTTACTGCGCAGTATCTCTATGAGGCTTTAA 1440  
 Db 1381 AATATCTGCTGCTTCTGAGAGAAATAGACTTACTGCGCAGTATCTCTATGAGGCTTTAA 1440  
 Oy 1441 TATGTAAGTGAAGAGTCACTCCAAATCAAGTATAGATATATTTGAAGCAAAATAT 1500  
 Db 1441 TATGTAAGTGAAGAGTCACTCCAAATCAAGTATAGATATATTTGAAGCAAAATAT 1500  
 Oy 1501 TTGGGAAACCTATCGGAAGAGAGGCAAGCCTCCCACTTAAGCCATGTAACCTGAATAATC 1560  
 Db 1501 TTGGGAAACCTATCGGAAGAGAGGCAAGCCTCCCACTTAAGCCATGTAACCTGAATAATC 1560



Db	1501	TTGGGAAACCTATTCGGAAAGGCAAGCCCTCCCACTTAAGCATTGTAATCTGAAATTC	1560
OY	1561	TAATTTATGAGACATTTTGGTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAATA	1620
Db	1561	TAATTTATGAGACATTTTGGTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAATA	1620
OY	1621	AATTTAAAGCCGTAAAGGAGACCTCATCAGGCGCTTCATCCGAGAGATTTTATCAAGAAAG	1680
Db	1621	AATTTAAAGCCGTAAAGGAGACCTCATCAGGCGCTTCATCCGAGAGATTTTATCAAGAAAG	1680
OY	1681	CAGATTTTGGCAGTTCCAAAAGACCTCGTGAATAAGATAATCAGGGAAGCTAACCAAGGAGC	1740
Db	1681	CAGATTTTGGCAGTTCCAAAAGACCTCGTGAATAAGATAATCAGGGAAGCTAACCAAGGAGC	1740
OY	1741	AGAATGGTCAAGTGTATTAATTTACTAATTAAGTGGTCATGAGATAATTAACAAGGTCATT	1800
Db	1741	AGAATGGTCAAGTGTATTAATTTACTAATTAAGTGGTCATGAGATAATTAACAAGGTCATT	1800
OY	1801	CTATTTAGCAATGAGAAAAAATCTTAACCCAAATAGATACCTCGAAAGAAATCTGCTTTCA	1860
Db	1801	CTATTTAGCAATGAGAAAAAATCTTAACCCAAATAGATACCTCGAAAGAAATCTGCTTTCA	1860
OY	1861	AAAGGAAAGCTGAACCTTAATAAGCAGCATATAGCAATATGGAATCGAATTAATATATCC	1920
Db	1861	AAAGGAAAGCTGAACCTTAATAAGCAGCATATAGCAATATGGAATCGAATTAATATATCC	1920
OY	1921	ACAATTTCAAAAGCACTTAATAAGAAATGGCTGAGAGGAAGTACTTTCAACCGGATATTC	1980
Db	1921	ACAATTTCAAAAGCACTTAATAAGAAATGGCTGAGAGGAAGTACTTTCAACCGGATATTC	1980
OY	1981	ATGCGCTTGAACCTAGTAGTAGTAAGAAATCTAAGCCCACTTAATTTGATAGTAATTTGCAA	2040
Db	1981	ATGCGCTTGAACCTAGTAGTAGTAAGAAATCTAAGCCCACTTAATTTGATAGTAATTTGCAA	2040
OY	2041	TTGATAGTGGTTCTAGCAGTGAAGATGAAGATTAAGAAAAAGTACAACCAATATGGCAGTCA	2100
Db	2041	TTGATAGTGGTTCTAGCAGTGAAGATGAAGATTAAGAAAAAGTACAACCAATATGGCAGTCA	2100
OY	2101	GGCAGCAGCAGAAACCTACAACTCATGTGAGGTAAGAACCTGCAACTGAGAGCCCAAGAA	2160
Db	2101	GGCAGCAGCAGAAACCTACAACTCATGTGAGGTAAGAACCTGCAACTGAGAGCCCAAGAA	2160
OY	2161	GTAACAGCCCAATGAGACAGACAAGTAAGAATGACAGTGATCTTTCCAGAGCTGA	2220
Db	2161	GTAACAGCCCAATGAGACAGACAAGTAAGAATGACAGTGATCTTTCCAGAGCTGA	2220
OY	2221	AGTTAACAAATGACACCTGGTCTTTTCTAGTAGTTCATAAATACCAAGTAATTAAGAT	2280
Db	2221	AGTTAACAAATGACACCTGGTCTTTTCTAGTAGTTCATAAATACCAAGTAATTAAGAT	2280
OY	2281	TTTGTCATCTAGACCTTCCAAAGAGAAAAAGAGAAATCTGAAGAACACTTAAGTGT	2340
Db	2281	TTTGTCATCTAGACCTTCCAAAGAGAAAAAGAGAAATCTGAAGAACACTTAAGTGT	2340
OY	2341	CTAATTAATGCTGGAAGACCCCAAAATCTCATGTTAAGTGGAGAAAGGTTTTCGAAACTG	2400
Db	2341	CTAATTAATGCTGGAAGACCCCAAAATCTCATGTTAAGTGGAGAAAGGTTTTCGAAACTG	2400
OY	2401	AAAATATCTGTAGAGATGACAGTATTTTCATTTGGTACTGAGTACGATTATGCGACTCAGG	2460
Db	2401	AAAATATCTGTAGAGATGACAGTATTTTCATTTGGTACTGAGTACGATTATGCGACTCAGG	2460
OY	2461	AAAGTATCTGCTTACTGGAAGTATGACACTYTAGGGAAGCAAAAACAGAACCAATTAAT	2520
Db	2461	AAAGTATCTGCTTACTGGAAGTATGACACTYTAGGGAAGCAAAAACAGAACCAATTAAT	2520
OY	2521	GTGAGAGTCACTGTGACAGATTTTAAAAACCCCAAGGGAGCTAATCATGGTGTGTTCCAAG	2580
Db	2521	GTGAGAGTCACTGTGACAGATTTTAAAAACCCCAAGGGAGCTAATCATGGTGTGTTCCAAG	2580
OY	2581	ATTAATAGAAATGACACAGAAAGCTTTAAGTATCCATTGGGACATGAAGTTAACCACAGTC	2640
Db	2581	ATTAATAGAAATGACACAGAAAGCTTTAAGTATCCATTGGGACATGAAGTTAACCACAGTC	2640

Db	2581	ATATATAGAAATATGACACACAAAGGCTTTAAGTATTCATTTGGACATGTAAGCTTAACCAACTC	2644.0
OY	2641	GGGAACAAGCATAGAAAAATGGAAGAAGTGAACCTTGATGCTCACTATTTTGCAGAAATACAT	2700.0
Db	2641	GGGAACAAGCATAGAAATGGAAGAAGTGAACCTTGATGCTCACTATTTTGCAGAAATACAT	2700.0
OY	2701	TCAAGGTTTCAAGGGCCAGTCACTTTGCTGCTTTTCAAAATCCAGGAATCCAGAAAGG	2766.0
Db	2701	TCAAGGTTTCAAGGGCCAGTCACTTTGCTGCTTTTCAAAATCCAGGAATCCAGAAAGG	2766.0
OY	2761	AATGTGCACACTTCTCTGCCCCACTCTGGGTCCTTAAAGAAACCAAGTCCAAAGTCACTT	2820.0
Db	2761	AATGTGCACACTTCTCTGCCCCACTCTGGGTCCTTAAAGAAACCAAGTCCAAAGTCACTT	2820.0
OY	2821	TTTGAATGTGAACAACAAAGGAAGAAATCAAGGAAGAATGAGTCTATATTCAGCCTGTAC	2880.0
Db	2821	TTTGAATGTGAACAACAAAGGAAGAAATCAAGGAAGAATGAGTCTATATTCAGCCTGTAC	2880.0
OY	2881	AGACAGTTAAATATCACTCGAGGCTTCTCTGAGTGGTGACAGAAAGATTAAGCCAGTTGATA	2940.0
Db	2881	AGACAGTTAAATATCACTCGAGGCTTCTCTGAGTGGTGACAGAAAGATTAAGCCAGTTGATA	2940.0
OY	2941	ATGCCAAATGTAGTATCAAAAGAGGCTCTAGGTTTGTCTATCAATCTCAGTTCAAGAGCA	3000.0
Db	2941	ATGCCAAATGTAGTATCAAAAGAGGCTCTAGGTTTGTCTATCAATCTCAGTTCAAGAGCA	3000.0
OY	3001	ACGAACCTGGAGCTACTTACTCCAAATTAACATGACACTTTTACAAACCCATATGCTATAC	3066.0
Db	3001	ACGAACCTGGAGCTACTTACTCCAAATTAACATGACACTTTTACAAACCCATATGCTATAC	3066.0
OY	3061	CACCACCTTTTCCCATCAAGTCACTTTTGTTTAAACTAAATGTATAGAAAAATCTGCTAGAG	3120.0
Db	3061	CACCACCTTTTCCCATCAAGTCACTTTTGTTTAAACTAAATGTATAGAAAAATCTGCTAGAG	3120.0
OY	3121	AAACCTTTGAGGAACATTCATGTCTCACTGAAAGAAATGGGAATGAGAACTTCCAA	3180.0
Db	3121	AAACCTTTGAGGAACATTCATGTCTCACTGAAAGAAATGGGAATGAGAACTTCCAA	3180.0
OY	3181	GTACAGTGTAGACAAATTAAGCCGTATATTAACATTAGAGAAATGTTTAAACAAACCCAGCT	3240.0
Db	3181	GTACAGTGTAGACAAATTAAGCCGTATATTAACATTAGAGAAATGTTTAAACAAACCCAGCT	3240.0
OY	3241	CAAGCATATTAATGAGTAGTCTCCAGTACTAATGAATGGGCTCCAGTATTAATGA	3300.0
Db	3241	CAAGCATATTAATGAGTAGTCTCCAGTACTAATGAATGGGCTCCAGTATTAATGA	3300.0
OY	3301	TAGGTTCCAGTATGAAACAACTTCACAGAGAACTAGTAGAAGAAAGAGGCCAAATTTGA	3366.0
Db	3301	TAGGTTCCAGTATGAAACAACTTCACAGAGAACTAGTAGAAGAAAGAGGCCAAATTTGA	3366.0
OY	3361	ATGCTATGCTTAGATTAGGGGTTTTTGCACCTGAGGTCATAAACAAGTCTTCTCGGAA	3420.0
Db	3361	ATGCTATGCTTAGATTAGGGGTTTTTGCACCTGAGGTCATAAACAAGTCTTCTCGGAA	3420.0
OY	3421	GTAATTTGTAAGCATCTCGAATAAATAAAGCAAGATATGAAGAGTAGTTAGACTGTA	3480.0
Db	3421	GTAATTTGTAAGCATCTCGAATAAATAAAGCAAGATATGAAGAGTAGTTAGACTGTA	3480.0
OY	3481	ATTACAGATTTCTCTCCATATCTGATTTTCAAGTAACCTTAGAACACCGTATGGGAAGTAGTC	3540.0
Db	3481	ATTACAGATTTCTCTCCATATCTGATTTTCAAGTAACCTTAGAACACCGTATGGGAAGTAGTC	3540.0
OY	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGCTGAATTAAGG	3600.0
Db	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGCTGAATTAAGG	3600.0
OY	3601	AAGTACTAGTTTGTCTGTAATAATACATTAAGGAAGTTCGCTGCTTTTAAACAAACCG	3660.0
Db	3601	AAGTACTAGTTTGTCTGTAATAATACATTAAGGAAGTTCGCTGCTTTTAAACAAACCG	3660.0
OY	3661	TCGAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCACATACATTTGGCTCAGG	3720.0
Db	3661	TCGAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCACATACATTTGGCTCAGG	3720.0

QY	3721	GTACCGAAGAGGGCCAAAGAAATTTAGAGTCCCTGAGAGAGAACTTATCTAGTGAGATG	3780
Db	3721	GTATCCGAAGAGGGGCCAAAGAAATTTAGAGTCCCTGAGAGAGAACTTATCTAGTGAGATG	3780
QY	3781	AAGAGCTTCCCTGCTTCCACACATTTGTTATTTGGTAAGTAACAAATATACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTTCCACACATTTGTTATTTGGTAAGTAACAAATATACCTTCTCAGT	3840
QY	3841	CTACTAGCATAGCAGCGTTGCTACCGAGTGTCTGTAAGACACAGAGGAAATTTAT	3900
Db	3841	CTACTAGCATAGCAGCGTTGCTACCGAGTGTCTGTAAGACACAGAGGAAATTTAT	3900
QY	3901	TATCATTTGAAGAAATAGCTTAAATGATGCTGCAACAGATATATTTGCCAAAGGCATCTC	3960
Db	3901	TATCATTTGAAGAAATAGCTTAAATGATGCTGCAACAGATATATTTGCCAAAGGCATCTC	3960
QY	3961	AGGAACATCAACCTTAGAGAGGAAACAAATGTTCTGCTAGTGTGTTTCTTCCACAGTCA	4020
Db	3961	AGGAACATCAACCTTAGAGAGGAAACAAATGTTCTGCTAGTGTGTTTCTTCCACAGTCA	4020
QY	4021	GTCATTTGAAGAACTTACTGCTCAATATCAAAACACCAGATCTTTCTTGATGGTCTT	4080
Db	4021	GTCATTTGAAGAACTTACTGCTCAATATCAAAACACCAGATCTTTCTTGATGGTCTT	4080
QY	4081	CCAAACAATGAGGCATCAGTCTGAAAGCCAGGAGTTGCTGAGTACAGAGAAATGG	4140
Db	4081	CCAAACAATGAGGCATCAGTCTGAAAGCCAGGAGTTGCTGAGTACAGAGAAATGG	4140
QY	4141	TTTCAGATGATGAAGAAAGAGGAGGCTTGGAAAGAAATATCAACAAGCAACAA	4200
Db	4141	TTTCAGATGATGAAGAAAGAGGAGGCTTGGAAAGAAATATCAACAAGCAACAA	4200
QY	4201	TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGAGAGTGAAGAAACAGCCTTCTGAG	4260
Db	4201	TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGAGAGTGAAGAAACAGCCTTCTGAG	4260
QY	4261	ACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAAACCATCTCAGAGAGGATACCATGC	4320
Db	4261	ACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAAACCATCTCAGAGAGGATACCATGC	4320
QY	4321	AACATTAACCTGATTAAGCTCCAGCAGAAATGCGTGAACCTGAAGCTGTGTAAGACAGC	4380
Db	4321	AACATTAACCTGATTAAGCTCCAGCAGAAATGCGTGAACCTGAAGCTGTGTAAGACAGC	4380
QY	4381	ATGGAGCCAGCCTTCTTAACAGCTAACCTTTCATCATAGTACTCTTCTGCCCTTGAGG	4440
Db	4381	ATGGAGCCAGCCTTCTTAACAGCTAACCTTTCATCATAGTACTCTTCTGCCCTTGAGG	4440
QY	4441	ACCTGCGAAATCCGAACAAAGCAGATCAGAAAAAGCAGTATTAATCTTCAAGAAAAATA	4500
Db	4441	ACCTGCGAAATCCGAACAAAGCAGATCAGAAAAAGCAGTATTAATCTTCAAGAAAAATA	4500
QY	4501	GTCATTAACCTATTAAGCCAGATCCAGAGGCTTTCTGCTGACAAAGTTTAAAGGCTGTG	4560
Db	4501	GTCATTAACCTATTAAGCCAGATCCAGAGGCTTTCTGCTGACAAAGTTTAAAGGCTGTG	4560
QY	4561	CAGATAGTTCTTACAGATAAATAAAGAACCCAGAGTGAAGAGTCAATCCCTTCTTAAT	4620
Db	4561	CAGATAGTTCTTACAGATAAATAAAGAACCCAGAGTGAAGAGTCAATCCCTTCTTAAT	4620
QY	4621	GCCCATCATTAAGATGATAGTGTGATGCAACAGTGTCTGAGAGTCTTCAAGATAGAA	4680
Db	4621	GCCCATCATTAAGATGATAGTGTGATGCAACAGTGTCTGAGAGTCTTCAAGATAGAA	4680
QY	4681	ACTAACCATCTCAAGAGAGCTCATTTAAGTTGTTGATGTGAGAGACAAAGCTGGAAG	4740
Db	4681	ACTAACCATCTCAAGAGAGCTCATTTAAGTTGTTGATGTGAGAGACAAAGCTGGAAG	4740
QY	4741	AGTCTGGGCCACAGATTTGACGGAACATCTTACTTGCCCAAGGCAAGATCTAAGAGGAA	4800
Db	4741	AGTCTGGGCCACAGATTTGACGGAACATCTTACTTGCCCAAGGCAAGATCTAAGAGGAA	4800
QY	4801	CCCTTACCTGGAATCTGGAATCAACCTCTTCTCTGATGACCCGTAATCTGATCCTTCTG	4860
Db	4801	CCCTTACCTGGAATCTGGAATCAACCTCTTCTCTGATGACCCGTAATCTGATCCTTCTG	4860
QY	4861	AAGACAGAGCCCCAGAGTCACTGTGTGGCAACATACATCTTCAACTCTGCAATTTGA	4920
Db	4861	AAGACAGAGCCCCAGAGTCACTGTGTGGCAACATACATCTTCAACTCTGCAATTTGA	4920
QY	4921	AAGTCCCAATTTGAAGTTGACAGATCTGCCAGAGTCCAGATGCTGCTCATACCTACTG	4980
Db	4921	AAGTCCCAATTTGAAGTTGACAGATCTGCCAGAGTCCAGATGCTGCTCATACCTACTG	4980
QY	4981	ATACTGCTGGTATTAATGCAATGGAAGAAAGTGTGACAGAGGAGAGCCAGAAATGACAG	5040
Db	4981	ATACTGCTGGTATTAATGCAATGGAAGAAAGTGTGACAGAGGAGAGCCAGAAATGACAG	5040
QY	5041	CTTCAACAGAAAGGCTCAACAAAGAAAGTCCATGAGTGTGTGAGCTTACCCCAAG	5100
Db	5041	CTTCAACAGAAAGGCTCAACAAAGAAAGTCCATGAGTGTGTGAGCTTACCCCAAG	5100
QY	5101	AATTTATGCTGCTGACAGTTTCCAGAAACACACATCACTTTAACTAATCTAATTA	5160
Db	5101	AATTTATGCTGCTGACAGTTTCCAGAAACACACATCACTTTAACTAATCTAATTA	5160
QY	5161	CTGAAGAGACTACATGATGTTATGAAGACAGATGCTGAGTGTGTGAACGAGAC	5220
Db	5161	CTGAAGAGACTACATGATGTTATGAAGACAGATGCTGAGTGTGTGAACGAGAC	5220
QY	5221	TGAATATATTTCTAGGAATTTGGGAGAGAAATGGTAGTATTCTTGGGTGACCC	5280
Db	5221	TGAATATATTTCTAGGAATTTGGGAGAGAAATGGTAGTATTCTTGGGTGACCC	5280
QY	5281	AGCTATTTAAAGAAAGAAATGCTGATGATGAGATTTTGAAGTACAGAGAGATGTG	5340
Db	5281	AGCTATTTAAAGAAAGAAATGCTGATGATGAGATTTTGAAGTACAGAGAGATGTG	5340
QY	5341	TCATGGAAGAACCCACCAAGGTCCTCAAGAGCAGAAAGATCCACAGAAAGATCT	5400
Db	5341	TCATGGAAGAACCCACCAAGGTCCTCAAGAGCAGAAAGATCCACAGAAAGATCT	5400
QY	5401	TCAGGGGCTGGAATGCTGTTATGGGCTTACCAACATGCCAGATCACTGTG	5460
Db	5401	TCAGGGGCTGGAATGCTGTTATGGGCTTACCAACATGCCAGATCACTGTG	5460
QY	5461	AATGATGATGACACTGTGTGCTCTGTGTGGAAGAGCTTTCATATTCACCTTG	5520
Db	5461	AATGATGATGACACTGTGTGCTCTGTGTGGAAGAGCTTTCATATTCACCTTG	5520
QY	5521	GCACAGGTGTCCACCCAAATTTGTGTGACACCCAGATGCCCTGACAGAGACAATGGCT	5580
Db	5521	GCACAGGTGTGTCCACCCAAATTTGTGTGACACCCAGATGCCCTGACAGAGACAATGGCT	5580
QY	5581	TCCATGCAATTTGGGACAGATGTGTGAGGACACTGTGTGACCCAGAGTGGGTTGGACA	5640
Db	5581	TCCATGCAATTTGGGACAGATGTGTGAGGACACTGTGTGACCCAGAGTGGGTTGGACA	5640
QY	5641	GTGTAGCACTTACAGTGTCCAGAGAGCTGACACATCTGATTAACCCAGATCCCCACA	5700
Db	5641	GTGTAGCACTTACAGTGTCCAGAGAGCTGACACATCTGATTAACCCAGATCCCCACA	5700
QY	5701	GCCACTACTGA 5711	
Db	5701	GCCACTACTGA 5711	
RESULT 11			
ARI36942			
LOCUS			
Sequence 1 from patent us 6162897.			
ACCESSION			
ARI36942.1 GI:14478192			
VERSION			
KEYWORDS			
Unknown.			



QY 1921 ACATTCACAAAGCAGCTTAAGAAATAGGCTGAGAGAGAGCTCTTCTACCGAGCATATTC 1980  
 Db 1921 ACATTCACAAAGCAGCTTAAGAAATAGGCTGAGAGAGAGCTCTTCTACCGAGCATATTC 1980  
 QY 1981 ATGGGCTGAGACTAGTGTAGTGAATCTAAGCCCACTAATGTCTGCAATTCGAAA 2040  
 Db 1981 ATGGGCTGAGACTAGTGTAGTGAATCTAAGCCCACTAATGTCTGCAATTCGAAA 2040  
 QY 2041 TTGATAGTTGTTTACGAGTGAAGAGATTAAGAAAAAAAGTACACCAATTCGAGTCA 2100  
 Db 2041 TTGATAGTTGTTTACGAGTGAAGAGATTAAGAAAAAAAGTACACCAATTCGAGTCA 2100  
 QY 2101 GGCACAGCAGAAACCTTCAACTCATGTGAAGTAAAGAACTGCAACTGAGCCAGAGA 2160  
 Db 2101 GGCACAGCAGAAACCTTCAACTCATGTGAAGTAAAGAACTGCAACTGAGCCAGAGA 2160  
 QY 2161 GTACACAGCCAAATGAGACAGACAGTAAAGACATGACAGTACTTCCAGAGCTGA 2220  
 Db 2161 GTACACAGCCAAATGAGACAGACAGTAAAGACATGACAGTACTTCCAGAGCTGA 2220  
 QY 2221 AGTTAACAAATGCACTGGTCTTTTACTAAGTGTCCAAATACGAGTAAAGAT 2280  
 Db 2221 AGTTAACAAATGCACTGGTCTTTTACTAAGTGTCCAAATACGAGTAAAGAT 2280  
 QY 2281 TTGTCAATCTAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
 Db 2281 TTGTCAATCTAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
 QY 2341 CTATATATGCTGAAG 2400  
 Db 2341 CTATATATGCTGAAG 2400  
 QY 2401 AAAGATCTGTAGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2460  
 Db 2401 AAAGATCTGTAGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2460  
 QY 2461 AAAGATCTGTAGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2520  
 Db 2461 AAAGATCTGTAGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2520  
 QY 2521 GTGTAGTCTAGTGTGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2580  
 Db 2521 GTGTAGTCTAGTGTGAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2580  
 QY 2581 ATATAGAAATGACACAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2640  
 Db 2581 ATATAGAAATGACACAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2640  
 QY 2641 GGGAAACACAGATTAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700  
 Db 2641 GGGAAACACAGATTAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700  
 QY 2701 TCAGAGTTTCAAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2760  
 Db 2701 TCAGAGTTTCAAGAGAGAGAGATTTTCAATGTTGACCTGATCTGATTTGCACTGAG 2760  
 QY 2761 AATGTGCAACATTTCTGCGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820  
 Db 2761 AATGTGCAACATTTCTGCGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820  
 QY 2821 TTGAAATGTGAACAAAG 2880  
 Db 2821 TTGAAATGTGAACAAAG 2880  
 QY 2881 AGACAGTTAATATACAG 2940  
 Db 2881 AGACAGTTAATATACAG 2940  
 QY 2941 ATGCCAAATGTAGATCAAG 3000  
 Db 2941 ATGCCAAATGTAGATCAAG 3000  
 QY 3001 ACGAAACTGAGCATTAATCTCAAAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060

Db 3001 ACGAAACTGAGCATTAATCTCAAAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060  
 QY 3061 CACCACTTTTCCATCAAGCATTTGTTAAAGTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
 Db 3061 CACCACTTTTCCATCAAGCATTTGTTAAAGTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
 QY 3121 AAACTTTGAG 3180  
 Db 3121 AAACTTTGAG 3180  
 QY 3181 GTACAGTGAAG 3240  
 Db 3181 GTACAGTGAAG 3240  
 QY 3241 CAAGCAATTAATGAAGTACAGTCCAGTCAATGAAGTGGGCTCAGATTAATGAAGAGAGAGAG 3300  
 Db 3241 CAAGCAATTAATGAAGTACAGTCCAGTCAATGAAGTGGGCTCAGATTAATGAAGAGAGAGAG 3300  
 QY 3301 TAGGTTCCAGTGAAG 3360  
 Db 3301 TAGGTTCCAGTGAAG 3360  
 QY 3361 ATGCTATGCTAGATTAAG 3420  
 Db 3361 ATGCTATGCTAGATTAAG 3420  
 QY 3421 GTAAATGTAACATCAAG 3480  
 Db 3421 GTAAATGTAACATCAAG 3480  
 QY 3481 ATACAGATTTCTCCATATCTGATTCAGTAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 Db 3481 ATACAGATTTCTCCATATCTGATTCAGTAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 QY 3541 ATGATCTCAGAGTTGTTTGAAG 3600  
 Db 3541 ATGATCTCAGAGTTGTTTGAAG 3600  
 QY 3601 AAGATACAGTTTGTGAG 3660  
 Db 3601 AAGATACAGTTTGTGAG 3660  
 QY 3661 TCAG 3720  
 Db 3661 TCAG 3720  
 QY 3721 GTTACCGAG 3780  
 Db 3721 GTTACCGAG 3780  
 QY 3781 AAGAGCTTCCCTGCTCCACAGCTTGTATTTGTAAGTAAACATATATCTCTCAGT 3840  
 Db 3781 AAGAGCTTCCCTGCTCCACAGCTTGTATTTGTAAGTAAACATATATCTCTCAGT 3840  
 QY 3841 CTACTAGGAGATGAG 3900  
 Db 3841 CTACTAGGAGATGAG 3900  
 QY 3901 TATCATTGAG 3960  
 Db 3901 TATCATTGAG 3960  
 QY 3961 AGGAAACATCACTTATGAG 4020  
 Db 3961 AGGAAACATCACTTATGAG 4020  
 QY 4021 GTGAATTTGAG 4080  
 Db 4021 GTGAATTTGAG 4080  
 QY 4081 CCAACAAATGAGAGATGAG 4140



Db 121 TGCATTTATCTGCTCTTCCGCTTGAGAGATACAAAATGTCATTATATCTATGCAAGAAA 180  
QY 181 TCTTAGAGTGTCCATCTGTCTGGAGTGTATCAAGAAACCTGTCTCCAAAGTGTGACC 240  
Db 181 TCTTAGAGTGTCCATCTGTCTGGAGTGTATCAAGAAACCTGTCTCCAAAGTGTGACC 240  
QY 241 ACATATTTTGCAAATTTTGCATGCTGAAACCTTCTCAACGACAGAAAGGCCCTTACAGT 300  
Db 241 ACATATTTTGCAAATTTTGCATGCTGAAACCTTCTCAACGACAGAAAGGCCCTTACAGT 300  
QY 301 GTCCCTTATGTAAGATGATATTAACCAAAAGAGCCCTCAAGAAAGTACGAGATTATATC 360  
Db 301 GTCCCTTATGTAAGATGATATTAACCAAAAGAGCCCTCAAGAAAGTACGAGATTATATC 360  
QY 361 AACTTGTGAAGACCTATTTGAAAAATCATTTTGTGCTTTTACGCTTGACACAGTTTGGAGT 420  
Db 361 AACTTGTGAAGACCTATTTGAAAAATCATTTTGTGCTTTTACGCTTGACACAGTTTGGAGT 420  
QY 421 ATGCAAAACAGCTATTAATTTTGCAAAAAAGAAAAATTAATCTCTGACATCTTAAAAAGT 480  
Db 421 ATGCAAAACAGCTATTAATTTTGCAAAAAAGAAAAATTAATCTCTGACATCTTAAAAAGT 480  
QY 481 AAGTTTATCATCCAAAGTATGGCTACAGAAACCCGTCAGAAAAAGACTTCTACAGAGT 540  
Db 481 AAGTTTATCATCCAAAGTATGGCTACAGAAACCCGTCAGAAAAAGACTTCTACAGAGT 540  
QY 541 AACCCGAAAAATCCCTCTTCTGAGAGAAACAGCTCAGTGTCCAACTCTTAACCTTGGAA 600  
Db 541 AACCCGAAAAATCCCTCTTCTGAGAGAAACAGCTCAGTGTCCAACTCTTAACCTTGGAA 600  
QY 601 CTGTGAGACTCTGAGAGCAAAAGCGGAGATACAACTCAAAAGAGCTGTGTACATTTG 660  
Db 601 CTGTGAGACTCTGAGAGCAAAAGCGGAGATACAACTCAAAAGAGCTGTGTACATTTG 660  
QY 661 AATTGGATGATCTTCTGAAAGATACCGTTAATAGGCAACTATTTGCACTGTGGAG 720  
Db 661 AATTGGATGATCTTCTGAAAGATACCGTTAATAGGCAACTATTTGCACTGTGGAG 720  
QY 721 ATCAAGAATTTGTAACAATCAACCCCTCAAGGAACAGGAGATGAATCAGTTTGGATTCTG 780  
Db 721 ATCAAGAATTTGTAACAATCAACCCCTCAAGGAACAGGAGATGAATCAGTTTGGATTCTG 780  
QY 781 CAAAAAAGGCTGCTGTGGAATTTTCTGAGAGGATGTAACAATATGMAATCATCATCAC 840  
Db 781 CAAAAAAGGCTGCTGTGGAATTTTCTGAGAGGATGTAACAATATGMAATCATCATCAC 840  
QY 841 CCAGTAAATATGATTTGAACACCACTGAGAGCGTGCAGCTGAGAGGATCCAGAAAGT 900  
Db 841 CCAGTAAATATGATTTGAACACCACTGAGAGCGTGCAGCTGAGAGGATCCAGAAAGT 900  
QY 901 ATCAGGGTACTTCTGTTTCAAACTTGCATGTGAGCCATGTGCAACAATCTCATGCA 960  
Db 901 ATCAGGGTACTTCTGTTTCAAACTTGCATGTGAGCCATGTGCAACAATCTCATGCA 960  
QY 961 GCTCATTACGCAATGAGAAACAGCAGTTTATTACTCACTAAAGACAGATGATAGAAA 1020  
Db 961 GCTCATTACGCAATGAGAAACAGCAGTTTATTACTCACTAAAGACAGATGATAGAAA 1020  
QY 961 GCTCATTACGCAATGAGAAACAGCAGTTTATTACTCACTAAAGACAGATGATAGAAA 1020  
Db 961 GCTCATTACGCAATGAGAAACAGCAGTTTATTACTCACTAAAGACAGATGATAGAAA 1020  
QY 1021 AGGCTGAATCTGTAATTAAGCAAAACAGCCTGCTTACGAGAGGCAACATTAACGAT 1080  
Db 1021 AGGCTGAATCTGTAATTAAGCAAAACAGCCTGCTTACGAGAGGCAACATTAACGAT 1080  
QY 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGGGAGCTCCAGACAGAAAAAGGTAG 1140  
Db 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGGGAGCTCCAGACAGAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAACCAAACTGCAATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAACCAAACTGCAATGCT 1200  
QY 1201 CAGAGATCCTTAAGATTAAGATGATGCTTGGATTAACACTTAATAGCAGCATTCAGA 1260  
Db 1201 CAGAGATCCTTAAGATTAAGATGATGCTTGGATTAACACTTAATAGCAGCATTCAGA 1260

QY 1261 AAGTTAATGATGTTTTCCAGAGTATGTAAGTGTAGTTGATGATGACTCACAGATG 1320  
Db 1261 AAGTTAATGATGTTTTCCAGAGTATGTAAGTGTAGTTGATGATGACTCACAGATG 1320  
QY 1321 GGGAGTCTGAATCAATATCCCAAGTACGTGATGATTGAGAGCTTCTAATAGGTAGATG 1380  
Db 1321 GGGAGTCTGAATCAATATCCCAAGTACGTGATGATTGAGAGCTTCTAATAGGTAGATG 1380  
QY 1381 AATATTCGTGTTCTTCAGAGAAATATAGCTTACGAGAGTACCTCATGAGGCTTTAA 1440  
Db 1381 AATATTCGTGTTCTTCAGAGAAATATAGCTTACGAGAGTACCTCATGAGGCTTTAA 1440  
QY 1441 TATGTAAGAAGTGAAGAGTTCTCTCAAAATCAGTAGAGATTAATTTGAAGACAAAATAT 1500  
Db 1441 TATGTAAGAAGTGAAGAGTTCTCTCAAAATCAGTAGAGATTAATTTGAAGACAAAATAT 1500  
QY 1501 TTGGGAAAACCTATCGAGAGAGCAAGCCCTCCCACTTAAGCATGTAACTGAAATTC 1560  
Db 1501 TTGGGAAAACCTATCGAGAGAGCAAGCCCTCCCACTTAAGCATGTAACTGAAATTC 1560  
QY 1561 TAATTTATAGAGCATTTTGTACTGAGCCACAGATTAATACAAAGCCCTCCACAAATA 1620  
Db 1561 TAATTTATAGAGCATTTTGTACTGAGCCACAGATTAATACAAAGCCCTCCACAAATA 1620  
QY 1621 AATTAAAGCGTAAAGAGAGACCTATCAGGCTTCTCTGAGAGATTTATCAAGAAAG 1680  
Db 1621 AATTAAAGCGTAAAGAGAGACCTATCAGGCTTCTCTGAGAGATTTATCAAGAAAG 1680  
QY 1681 CAGATTTTGGCAATTCAAAAGACTCTGTAATATGATTAATCAAGGAACTTAACCAAGGAGC 1740  
Db 1681 CAGATTTTGGCAATTCAAAAGACTCTGTAATATGATTAATCAAGGAACTTAACCAAGGAGC 1740  
QY 1741 AGAATGGCAAGCATGAATTAATTAATAGTGTCTGATGATTAATTAACAAAGGATTT 1800  
Db 1741 AGAATGGCAAGCATGAATTAATTAATAGTGTCTGATGATTAATTAACAAAGGATTT 1800  
QY 1801 CTATTGAGAAATGAGAAAAATCTTAACCAATAGAAATCTACGAAAAAAGAAATCTGTTTCA 1860  
Db 1801 CTATTGAGAAATGAGAAAAATCTTAACCAATAGAAATCTACGAAAAAAGAAATCTGTTTCA 1860  
QY 1861 AAACGAAAAGCTGAACCTATTAAGCAGCAGTATTAAGCAATATGCAATTAATATATCC 1920  
Db 1861 AAACGAAAAGCTGAACCTATTAAGCAGCAGTATTAAGCAATATGCAATTAATATATCC 1920  
QY 1921 ACAATTCAAAAAGCCTTAATAAAGATAGGCTGAGAGGAGAGTCTTACCAGGCAATTC 1980  
Db 1921 ACAATTCAAAAAGCCTTAATAAAGATAGGCTGAGAGGAGAGTCTTACCAGGCAATTC 1980  
QY 1981 ATGGGCTTGAACCTAGTATGATGAATCTAAGCCACCTAATTTAGTGAATTTGCAAA 2040  
Db 1981 ATGGGCTTGAACCTAGTATGATGAATCTAAGCCACCTAATTTAGTGAATTTGCAAA 2040  
QY 2041 TTGATAGTGTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTCAACCAATGCGCAGTCA 2100  
Db 2041 TTGATAGTGTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTCAACCAATGCGCAGTCA 2100  
QY 2101 GGCACAGCAGAAACCTCAACTCATGGAAGGTAAAGACCTGCACTGAGGAGCCAGAAAGA 2160  
Db 2101 GGCACAGCAGAAACCTCAACTCATGGAAGGTAAAGACCTGCACTGAGGAGCCAGAAAGA 2160  
QY 2161 GTAAACAAGCCAAATGAACAGACAAATGAAGAGACATGACAGTACTTTCCAGAGCTGA 2220  
Db 2161 GTAAACAAGCCAAATGAACAGACAAATGAAGAGACATGACAGTACTTTCCAGAGCTGA 2220  
QY 2221 AGTTAACAATGCACTGGTTCTTTTACTAAGTGTTCAAATACACAGTAAAGAT 2280  
Db 2221 AGTTAACAATGCACTGGTTCTTTTACTAAGTGTTCAAATACACAGTAAAGAT 2280  
QY 2281 TTGTCAATCTAGCTTCCAGAGAGAAAGAAAGAAAGAAAGAAAGTAAAGTGT 2340  
Db 2281 TTGTCAATCTAGCTTCCAGAGAGAAAGAAAGAAAGAAAGAAAGTAAAGTGT 2340

QY 2341 CTAATATGCTGAGAACCCCAAGATCTCATGTAAAGTGAGAGAGGGTTTTCACAACTG 2400  
Db 2341 CTAATATGCTGAGAACCCCAAGATCTCATGTAAAGTGAGAGAGGGTTTTCACAACTG 2400  
QY 2401 AAAGATCTGTAGAGAGTAGACATATTTCATTGTGTACCTGTACTATTAATGTGCACTCAG 2460  
Db 2401 AAAGATCTGTAGAGAGTAGACATATTTCATTGTGTACCTGTACTATTAATGTGCACTCAG 2460  
QY 2461 AAAGATCTGTGTAGAGAGTAGACATCTTAGGAGAGGCAAAACAGAACCAATTAAT 2520  
Db 2461 AAAGATCTGTGTAGAGAGTAGACATCTTAGGAGAGGCAAAACAGAACCAATTAAT 2520  
QY 2521 GTGTAGTCAAGTGTGACCAATTTGAAACCCCAAGGACTAATTCATGTGTTCCAAAG 2580  
Db 2521 GTGTAGTCAAGTGTGACCAATTTGAAACCCCAAGGACTAATTCATGTGTTCCAAAG 2580  
QY 2581 ATAATAGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATAGTTAACCCAGTC 2640  
Db 2581 ATAATAGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATAGTTAACCCAGTC 2640  
QY 2641 GGAATACAGCATTAAGAAATGAGAAAGTGAACCTGATGCTGATTTGAGATACAT 2700  
Db 2641 GGAATACAGCATTAAGAAATGAGAAAGTGAACCTGATGCTGATTTGAGATACAT 2700  
QY 2701 TCAAGGTTTCAAGGCGCAGTCATTGCTGTGTTTCAATCCAGAAATGCAGAAAG 2760  
Db 2701 TCAAGGTTTCAAGGCGCAGTCATTGCTGTGTTTCAATCCAGAAATGCAGAAAG 2760  
QY 2761 AATGTGCAACATTTCTGCCCCCTGTGGTCTTAAAGAACAAAGTCCAAAGTCACTT 2820  
Db 2761 AATGTGCAACATTTCTGCCCCCTGTGGTCTTAAAGAACAAAGTCCAAAGTCACTT 2820  
QY 2821 TTGAATGTGACAAAGAGAAATCAAGAGAAATGAGTCTATATTCAGACCTGTAC 2880  
Db 2821 TTGAATGTGACAAAGAGAAATCAAGAGAAATGAGTCTATATTCAGACCTGTAC 2880  
QY 2881 AGACAGTTAATATCATCTGACGCTTCTGTGTTGTCAGAAAGATAGCCAGTTGATA 2940  
Db 2881 AGACAGTTAATATCATCTGACGCTTCTGTGTTGTCAGAAAGATAGCCAGTTGATA 2940  
QY 2941 ATGCCAATGTATATCAAGAGAGGCTAGGTTTGTCTATCATCTCAGTTCAAGGCA 3000  
Db 2941 ATGCCAATGTATATCAAGAGAGGCTAGGTTTGTCTATCATCTCAGTTCAAGGCA 3000  
QY 3001 AGCAATGAGCACTCATTAATTAACATGACTTTTACAAACCATTCGTATAC 3060  
Db 3001 AGCAATGAGCACTCATTAATTAACATGACTTTTACAAACCATTCGTATAC 3060  
QY 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAAGAAATCTGCTAGAG 3120  
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAAGAAATCTGCTAGAG 3120  
QY 3121 AAAAATTTGAGAACATTCATGTCACCTGAAAGAGAAATGGAATAGAAACATTTCAA 3180  
Db 3121 AAAAATTTGAGAACATTCATGTCACCTGAAAGAGAAATGGAATAGAAACATTTCAA 3180  
QY 3181 GTACAGTACGCAATTAAGCCTATTAACATTAGAGAAATGTTTAAAGAGGACAGT 3240  
Db 3181 GTACAGTACGCAATTAAGCCTATTAACATTAGAGAAATGTTTAAAGAGGACAGT 3240  
QY 3241 CAAGCAATTAATAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAGAA 3300  
Db 3241 CAAGCAATTAATAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAGAA 3300  
QY 3301 TAGGTTCCAGTATTAAGAAACATTCAGAGAACTAGTGAAGAGAGGCCCCAAATTTGA 3360  
Db 3301 TAGGTTCCAGTATTAAGAAACATTCAGAGAACTAGTGAAGAGAGGCCCCAAATTTGA 3360  
QY 3361 ATGCTATGCTTAGTAGGGGTTTTCAGAACCTGAGGCTATTAAGAAAGGCTTCTCTGGA 3420  
Db 3361 ATGCTATGCTTAGTAGGGGTTTTCAGAACCTGAGGCTATTAAGAAAGGCTTCTCTGGA 3420  
QY 3421 GTAAATGTAGACATCTGAAATAAAAAGCAAGATATGAAGAGTACAGCTGTTA 3480

Db 3421 GTAAATGTAGACATCTGAAATAAAAAGCAGAAATATGAAGAGTAGTTACAGCTGTA 3480  
QY 3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAACCTTAGAAGAGCTATGGAGTAGTC 3540  
Db 3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAACCTTAGAAGAGCTATGGAGTAGTC 3540  
QY 3541 ATGCAATCAGGTTTGTCTGAGACACCGTACCTGTATGATATGATGTTGAATTAAGG 3600  
Db 3541 ATGCAATCAGGTTTGTCTGAGACACCGTACCTGTATGATATGATGTTGAATTAAGG 3600  
QY 3601 AAGATACAGTTTCTGTAAGATTAAGAAAGTTCCTGTTTTCACAAAGCG 3660  
Db 3601 AAGATACAGTTTCTGTAAGATTAAGAAAGTTCCTGTTTTCACAAAGCG 3660  
QY 3661 TCCAGAAAGGAGGCTTAGCAGAGAGTCCAGCCCTTACCCATACACATTTGGCTCAG 3720  
Db 3661 TCCAGAAAGGAGGCTTAGCAGAGAGTCCAGCCCTTACCCATACACATTTGGCTCAG 3720  
QY 3721 GTTACCGAAGAGGCGCCAAAGAAATTAAGTCTCAGAGAGAACTATCTAGAGATG 3780  
Db 3721 GTTACCGAAGAGGCGCCAAAGAAATTAAGTCTCAGAGAGAACTATCTAGAGATG 3780  
QY 3781 AAGACCTTCCCTGCTTCCAAACACTGTTATTTGTAAGTAAACAATATACCTCTCAGT 3840  
Db 3781 AAGACCTTCCCTGCTTCCAAACACTGTTATTTGTAAGTAAACAATATACCTCTCAGT 3840  
QY 3841 CTACTAGGCTATGACACCGTGTCTACCGAGTGTCTGTAAAGAACACAGAGGAATTTAT 3900  
Db 3841 CTACTAGGCTATGACACCGTGTCTACCGAGTGTCTGTAAAGAACACAGAGGAATTTAT 3900  
QY 3901 TATCATTTGAGAAATGCTTAATTAATGACTGAGTAACAGAGTATATTGGCAAGGCAATC 3960  
Db 3901 TATCATTTGAGAAATGCTTAATTAATGACTGAGTAACAGAGTATATTGGCAAGGCAATC 3960  
QY 3961 AGGAATCATCACCTTAGTGAAGAGAAACAAATGTTGCTAGCTGTTTCTTCCACAGTGA 4020  
Db 3961 AGGAATCATCACCTTAGTGAAGAGAAACAAATGTTGCTAGCTGTTTCTTCCACAGTGA 4020  
QY 4021 GTGAATTTGAGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
Db 4021 GTGAATTTGAGAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
QY 4081 CCAACCAATGAGGATCAGTCTGAAGGAGGAGGAGTGTGAGTGAAGCAAGAAATGG 4140  
Db 4081 CCAACCAATGAGGATCAGTCTGAAGGAGGAGGAGTGTGAGTGAAGCAAGAAATGG 4140  
QY 4141 TTTGAGATGATGAGAAAGAGAGAGGAGGCTTGAAGAAATTAATCAAGAGAGCAAGCA 4200  
Db 4141 TTTGAGATGATGAGAAAGAGAGAGGAGGCTTGAAGAAATTAATCAAGAGAGCAAGCA 4200  
QY 4201 TGGATTCAAACTTAGTGAAGAGGATCTGGGTGAGAGTGAAGCAAGCCGCTGTGAAG 4260  
Db 4201 TGGATTCAAACTTAGTGAAGAGGATCTGGGTGAGAGTGAAGCAAGCCGCTGTGAAG 4260  
QY 4261 ACTGCTCAGGCTATCTCTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4320  
Db 4261 ACTGCTCAGGCTATCTCTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4320  
QY 4321 AACATTAACCTGATAAGCTCCAGAGAAATGGCTGAAGTGAAGTGAAGTGAAGTGAAG 4380  
Db 4321 AACATTAACCTGATAAGCTCCAGAGAAATGGCTGAAGTGAAGTGAAGTGAAGTGAAG 4380  
QY 4381 ATGGAGGAGGAGGCTTCTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4440  
Db 4381 ATGGAGGAGGAGGCTTCTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4440  
QY 4441 AACTGCAAAATCCAAACAAAGCAATGAGAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 4500  
Db 4441 AACTGCAAAATCCAAACAAAGCAATGAGAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 4500  
QY 4501 GTGAATTAACCTATAAGCAGAAATCCAGAGAGGCTTGTGCTGACAAAGTTGAGGTGTG 4560  
Db 4501 GTGAATTAACCTATAAGCAGAAATCCAGAGAGGCTTGTGCTGACAAAGTTGAGGTGTG 4560



Db	4501	GTGAATACCCTATTAAGCCAAAGTCCAGAAGCCCTTTCTGCTGACAAAGTTTAGGTGTCGTG	4580
OY	4561	CAGATAGTTCTPACCAAGTAAAAATAAAGAACACAGAGGTGGAAAGTCATCCCTCTCTAAAT	4620
Db	4561	CAGATAGTTCTPACCAAGTAAAAATAAAGAACACAGAGGTGGAAAGTCATCCCTCTCTAAAT	4620
OY	4621	GCCCATTCATTAGATGATAGGTGTGTATGCAATGCAACAGTGTCTCTGGGAGTCTTAGAAATAGAA	4680
Db	4621	GCCCATTCATTAGATGATAGGTGTGTATGCAATGCAACAGTGTCTCTGGGAGTCTTAGAAATAGAA	4680
OY	4681	ACTAACCCCTCCAGAGAGAGAGCTCATTTAAGTGTGTGATGTGGAGAGCAACAGCTGGAG	4740
Db	4681	ACTAACCCCTCCAGAGAGAGAGCTCATTTAAGTGTGTGATGTGGAGAGCAACAGCTGGAG	4740
OY	4741	AGTCTGGGCGACACGATTTTGAACGGAAACATCTTACTTGCCAAAGCCAAAGATCTAGAGGAA	4800
Db	4741	AGTCTGGGCGCGACGATTTTGAACGGAAACATCTTACTTGCCAAAGCCAAAGATCTAGAGGAA	4800
OY	4801	CCCCCTTACCTGGAAATCTGGAAATCAAGCCCTCTCTGATGACCCCTGAATCTGATCCTTCTG	4860
Db	4801	CCCCCTTACCTGGAAATCTGGAAATCAAGCCCTCTCTGATGACCCCTGAATCTGATCCTTCTG	4860
OY	4861	AAGACAGAGCCCCACAGAGTCAAGCTGTGTGGCAACATACATCTTCAACCTCCTGATTTGA	4920
Db	4861	AAGACAGAGCCCCACAGAGTCAAGCTGTGTGGCAACATACATCTTCAACCTCCTGATTTGA	4920
OY	4921	AAGTTCCCAATTTGAAGTTGACAGATCTGCCACAGATCCAGCTGCTGCTCATCTACTG	4980
Db	4921	AAGTTCCCAATTTGAAGTTGACAGATCTGCCACAGATCCAGCTGCTGCTCATCTACTG	4980
OY	4981	ATACTGCTGGGTATTAATCAATTCGAAAGAAATGTGTAGCAGGAGCAAGCCAGAAATTGACAG	5040
Db	4981	ATACTGCTGGGTATTAATCAATTCGAAAGAAATGTGTAGCAGGAGCAAGCCAGAAATTGACAG	5040
OY	5041	CTTCAACAGAAAGGCTCAACAAAAGAAATGTCATGATGTGTCTGGCTGACCCCAAG	5100
Db	5041	CTTCAACAGAAAGGCTCAACAAAAGAAATGTCATGATGTGTCTGGCTGACCCCAAG	5100
OY	5101	AATTTATGCTGCTGTACAAAGTTTGCCAGAAAACACACATCAGCTTAACTAATCTAAATTA	5160
Db	5101	AATTTATGCTGCTGTACAAAGTTTGCCAGAAAACACACATCAGCTTAACTAATCTAAATTA	5160
OY	5161	CTGAAGAGACATCACTGCTGTTTATGAAACACAGATGCTGAGTTTGTGTGTGAACGGACAC	5220
Db	5161	CTGAAGAGACATCACTGCTGTTTATGAAACACAGATGCTGAGTTTGTGTGTGAACGGACAC	5220
OY	5221	TGAAATATATTTCTAGGAATTCGGGAGAGAAATGGGTAGTATTTCTGGGTGACCC	5280
Db	5221	TGAAATATATTTCTAGGAATTCGGGAGAGAAATGGGTAGTATTTCTGGGTGACCC	5280
OY	5281	AGTCTATTTAAAGAAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTGG	5340
Db	5281	AGTCTATTTAAAGAAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTGG	5340
OY	5341	TCATTCGAAAGAAACCCACAGGTGCCAAAGCAGCAAGAGAATCCACAGACAGAAAGATCT	5400
Db	5341	TCATTCGAAAGAAACCCACAGGTGCCAAAGCAGCAAGAGAATCCACAGACAGAAAGATCT	5400
OY	5401	TCGAGGGGCTGGAATCTGTTGATATGGGCCCTTCCACCAACATGCTCCACAGATCAACTGG	5460
Db	5401	TCGAGGGGCTGGAATCTGTTGATATGGGCCCTTCCACCAACATGCTCCACAGATCAACTGG	5460
OY	5461	AATGAGATGGTACACGCTGTGTGCTCTCTGTGTGAAGAGAGCTTTCATCATTCACCTTG	5520
Db	5461	AATGAGATGGTACACGCTGTGTGCTCTCTGTGTGAAGAGAGCTTTCATCATTCACCTTG	5520
OY	5521	GCACAGGTGTCCACCCAAATTTGGTTGTGACGCGAGATGCTTGACAGAGACAAATGCT	5580
Db	5521	GCACAGGTGTCCACCCAAATTTGGTTGTGACGCGAGATGCTTGACAGAGACAAATGCT	5580
OY	5581	TTCATTCGAATTTGGGCGAATGTGTGTAGAGGCACTGTGTGTGACCCGAGAGTGGGTGTTGGACA	5640
Db	5581	TTCATTCGAATTTGGGCGAATGTGTGTAGAGGCACTGTGTGTGACCCGAGAGTGGGTGTTGGACA	5640

Query	Match	Best Local Similarity	Score	DB	Length
Query Match	99.9%	Score 5707.8	DB 6	Length 5914	
Best Local Similarity	100.0%	Pred. No. 0			
Matches 5709	Conservative 0	Mismatches 2	Indels 0	Gaps 0	
LOCUS	180938	5914 bp	DNA	Linear	PAT 10-JUN-1998
DEFINITION	Sequence 1 from patent US 5709999.				
ACCESSION	180938				
VERSION	180938.1	GI:3209228			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 5914)				
AUTHORS	Shattuck-Eidens,D.M., Simard,J., Durocher,F., Emi,M. and Nakamura,Y.				
TITLE	Linked breast and ovarian cancer susceptibility gene				
JOURNAL	Patent: US 5709999-A 1 20-JAN-1998;				
FEATURES	Location/Qualifiers				
Source	1..5914				
BASE COUNT	2006 a 1156 c 1316 g 1436 t				
ORIGIN					
1	AGCTGCTGAGACTTCGTGGAGCCCGCACACAGGCTGTGGGTTTCTCAGATAACTGGGCC	60			
2	AGCTGCTGAGACTTCGTGGAGCCCGCACACAGGCTGTGGGTTTCTCAGATAACTGGGCC	60			
3	CCTGGCTCAGAGAGGCTTACACCTCTCTCTGGGTAAGTTCATGGAAACAGAAAGAA	120			
4	CCTGGCTCAGAGAGGCTTACACCTCTCTCTGGGTAAGTTCATGGAAACAGAAAGAA	120			
5	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
6	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
7	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
8	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
9	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
10	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
11	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
12	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
13	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
14	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
15	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
16	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
17	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
18	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
19	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
20	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
21	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
22	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
23	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
24	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
25	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
26	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
27	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
28	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
29	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
30	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
31	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
32	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
33	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
34	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
35	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
36	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
37	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
38	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
39	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
40	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
41	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATGCTATGCGAGAAA	180			
42	TGGATTATCTGCTTTCGCGGTGAAGAAGTACAAATGCTAATG				

RESULT	13			
LOCUS	180938			
DEFINITION	180938	5914 bp	DNA	linear
ACCESSION	180938	Sequence 1 from patent US 5709399.		PAT 10-JUN-1998
VERSION	180938..1	GI:3209228		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 5914)			
	Shattuck-Eidens, D.M., Simard, J., Durocher, F., Emi, M. and Nakamura, Y.			
TITLE	Linked breast and ovarian cancer susceptibility gene			
JOURNAL	Patent: US 5709399-A 1 20-JAN-1998;			
FEATURES	Location/Qualifiers			
source	1..5914			
	/organism="unknown"			
BASE COUNT	2066 a	1156 c	1316 g	1436 t
ORIGIN				



Db 541 AACCCGAAATCTTCCTCGCAGGAAACAGTCTCAGTGTCTCACTCTCTAACCCTTGAA 600  
OY 601 CTGTGAGAACTCTGAGACAAAGACGGGATACAACTCAAAAAGAGCTGTGTACATG 660  
Db 601 CTGTGAGAACTCTGAGACAAAGACGGGATACAACTCAAAAAGAGCTGTGTACATG 660  
OY 661 AATGGAGTCTGATCTCTCTGAGATACCGTTAATAGGCAACCTTATTGCACTGGGAG 720  
Db 661 AATGGAGTCTGATCTCTCTGAGATACCGTTAATAGGCAACCTTATTGCACTGGGAG 720  
OY 721 ATCAGAAATGTACAAATACCCCTCAAGAACGAGGATGAATCACTTGGATTCG 780  
Db 721 ATCAGAAATGTACAAATACCCCTCAAGAACGAGGATGAATCACTTGGATTCG 780  
OY 781 CAAAAAGGCTCTGTGATTTCTGAGACGATGTAAACAATCTGAATCATCATC 840  
Db 781 CAAAAAGGCTCTGTGATTTCTGAGACGATGTAAACAATCTGAATCATCATC 840  
OY 841 CCAGTAATATGATTTGAAACACCACTGAGAGCGTGCAGCTGAGAGCCATCCAGAAAGT 900  
Db 841 CCAGTAATATGATTTGAAACACCACTGAGAGCGTGCAGCTGAGAGCCATCCAGAAAGT 900  
OY 901 ATCAGAGTATCTCTCTCAACCTTGATGATGAGACCATGTGCAACATCTCATGCCA 960  
Db 901 ATCAGAGTATCTCTCTCAACCTTGATGATGAGACCATGTGCAACATCTCATGCCA 960  
OY 961 GCTCATTCACGATGAGAACACGACGATTTATTACTACTAAGACAGAAATGATAGAA 1020  
Db 961 GCTCATTCACGATGAGAACACGACGATTTATTACTACTAAGACAGAAATGATAGAA 1020  
OY 1021 AGGCTGAATTTCTGTATAAAGCAACAGCCCTGGCTTAGCAGAGGCCAATTAACAGAT 1080  
Db 1021 AGGCTGAATTTCTGTATAAAGCAACAGCCCTGGCTTAGCAGAGGCCAATTAACAGAT 1080  
OY 1081 GGGCTGGAAGTGAAGAAACATGTATGATGAGCGGACTCCGACACAGAAAAAGGTAG 1140  
Db 1081 GGGCTGGAAGTGAAGAAACATGTATGATGAGCGGACTCCGACACAGAAAAAGGTAG 1140  
OY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGAGAGAACTGCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGAGAGAACTGCCATGCT 1200  
OY 1201 CAGAGATCTAGAGATACCTGAAGATGTTCTTGATTAACACTAAATAGCAGCATTCAGA 1260  
Db 1201 CAGAGATCTAGAGATACCTGAAGATGTTCTTGATTAACACTAAATAGCAGCATTCAGA 1260  
OY 1261 AAGTTAATGAGTGTTCCTGAGAAAGTGAATGTTAGTTCATGATGACATGATG 1320  
Db 1261 AAGTTAATGAGTGTTCCTGAGAAAGTGAATGTTAGTTCATGATGACATGATG 1320  
OY 1321 GGGAGTCTGAATCAATGCGAAAGTGAATGATGATGAGAGTTCCTAATGAGTAAAG 1380  
Db 1321 GGGAGTCTGAATCAATGCGAAAGTGAATGATGATGAGAGTTCCTAATGAGTAAAG 1380  
OY 1381 AATATTTCTGTTCTTCAGAGAAATAGACTTACTGGCAGTGAATCTCATGAGGCTTAA 1440  
Db 1381 AATATTTCTGTTCTTCAGAGAAATAGACTTACTGGCAGTGAATCTCATGAGGCTTAA 1440  
OY 1441 TATGTAAAAGTGAAGAGTTCCTCAATCAGTAGAGATTAATTTAAGCAAAATAT 1500  
Db 1441 TATGTAAAAGTGAAGAGTTCCTCAATCAGTAGAGATTAATTTAAGCAAAATAT 1500  
OY 1501 TTGGGAAAAACCTATGCGAAGAGGCAAGCTCCCAACTTAAGCCATGATGAAAAATC 1560  
Db 1501 TTGGGAAAAACCTATGCGAAGAGGCAAGCTCCCAACTTAAGCCATGATGAAAAATC 1560  
OY 1561 TAATTTATAGGACATTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCACAAATA 1620  
Db 1561 TAATTTATAGGACATTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCACAAATA 1620  
OY 1621 AATTAAGCGTAAAGAGAGCTACATGAGGCTTCATCTGAGAGATTTTATCAAGAAAG 1680  
Db 1621 AATTAAGCGTAAAGAGAGCTACATGAGGCTTCATCTGAGAGATTTTATCAAGAAAG 1680

OY 1681 CAGATTTGGCAGTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTAACCAAGGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTAACCAAGGAGC 1740  
OY 1741 AGAATGTCAGTGAATTAATTAATAGTGTATGAGATTAATTAAGAAAGTGAT 1800  
Db 1741 AGAATGTCAGTGAATTAATTAATAGTGTATGAGATTAATTAAGAAAGTGAT 1800  
OY 1801 CTATTCAGATGAGAAAAATCCTAACCCATAGATCACTGCAAAAAGAACTGCTTCA 1860  
Db 1801 CTATTCAGATGAGAAAAATCCTAACCCATAGATCACTGCAAAAAGAACTGCTTCA 1860  
OY 1861 AAACGAAAGCTTAACCTATTAAGCAGCAGTATAGCAATTAAGCACTGAATTAATATC 1920  
Db 1861 AAACGAAAGCTTAACCTATTAAGCAGCAGTATAGCAATTAAGCACTGAATTAATATC 1920  
OY 1921 ACAATTCAAAAGCACCTTAATAAAGATAGGCTGAGAGAGAGTCTTACAGGCATATTC 1980  
Db 1921 ACAATTCAAAAGCACCTTAATAAAGATAGGCTGAGAGAGAGTCTTACAGGCATATTC 1980  
OY 1981 ATGGCTTGAATAGTACGATAGCAATCTAAGCCACCTAATGTACTGAATTCGAA 2040  
Db 1981 ATGGCTTGAATAGTACGATAGCAATCTAAGCCACCTAATGTACTGAATTCGAA 2040  
OY 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTAAACCAATGCGCAGTCA 2100  
Db 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTAAACCAATGCGCAGTCA 2100  
OY 2101 GGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACTCACTGAGCCAGAAAGA 2160  
Db 2101 GGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACTCACTGAGCCAGAAAGA 2160  
OY 2161 GTAAACGCAATTAAGACAGACAACTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
Db 2161 GTAAACGCAATTAAGACAGACAACTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
OY 2221 AGTTAAACAAATGACCTGCTTCTTACTAATGTTCAAAATACAGTGAATTAAGAAAT 2280  
Db 2221 AGTTAAACAAATGACCTGCTTCTTACTAATGTTCAAAATACAGTGAATTAAGAAAT 2280  
OY 2281 TTGTCAATCTAGCCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGTGT 2340  
Db 2281 TTGTCAATCTAGCCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGTGT 2340  
OY 2341 CTAAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTGCAGAACTG 2400  
Db 2341 CTAAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTGCAGAACTG 2400  
OY 2401 AAAGATCTGAGAGTGAAGTATTTCAATGCTGCTGCTGCTGATTAATGACACTCAGG 2460  
Db 2401 AAAGATCTGAGAGTGAAGTATTTCAATGCTGCTGCTGCTGATTAATGACACTCAGG 2460  
OY 2461 AAAGATCTGCTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAATTAAT 2520  
Db 2461 AAAGATCTGCTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAATTAAT 2520  
OY 2521 GTGTAGTCAAGTGTCTAGCATTGTGAAGCCCAAGGACTAATTTATGTTTCCAAAG 2580  
Db 2521 GTGTAGTCAAGTGTCTAGCATTGTGAAGCCCAAGGACTAATTTATGTTTCCAAAG 2580  
OY 2581 ATAAATGAAGTGAACAGAGGCTTAAAGTATCCATTGGGACATGAAGTTAACCAAGT 2640  
Db 2581 ATAAATGAAGTGAACAGAGGCTTAAAGTATCCATTGGGACATGAAGTTAACCAAGT 2640  
OY 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAATGATGCTCAGTATTTGCGAATACAT 2700  
Db 2641 GGGAAACAAGCATAGAAATGGAAGAAAGTGAATGATGCTCAGTATTTGCGAATACAT 2700  
OY 2701 TCAAGTTTCAAGGCGCCAGTATTTGCTGTGTTTCAAAATCAGAGAAATGCAAGAGG 2760  
Db 2701 TCAAGTTTCAAGGCGCCAGTATTTGCTGTGTTTCAAAATCAGAGAAATGCAAGAGG 2760

QY	2761	AATGTCACACATTTCTCTGGCCACTCTGGGCTCTTAAGAACAAGTCCAAAAGTCACTT	2820
Db	2761		
QY	2821	TTGAATGTGAACAAAAGAAATCAAGGAAGAGTCAATATATCAAGCTGTAC	2880
Db	2821		
QY	2881	AGACAGTTAATATCACTGCAGGCTTTCGTGGTGGTCAGAAAGATTAAGCCAGTTGAT	2940
Db	2881		
QY	2941	ATGCCAAATGTAGTATCAAGAAGGCTCTAGTTTTGTCTATCATCTCAGTTGAGGCA	3000
Db	2941		
QY	3001	ACGAACCTGACATCTACTCCAAATTAACATGACTTTTACAAACCCATATCGATAC	3060
Db	3001		
QY	3061	CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTAAAGAAATCTGCTAGAG	3120
Db	3061		
QY	3121	AAACTTTGAGAACATTTAATGTCACTCCGTAAGAGAAATGGAAATGAGACATTTCCA	3180
Db	3121		
QY	3181	GTACAGTGAACAATTAAGCCGTAATACATTAGAGAAATGTTTTAAAGAACGACGT	3240
Db	3181		
QY	3241	CAGCAATATTAATGAAGTAGTTCCTCACTAATGAAGTGGGCTCCAGTATTATGA	3300
Db	3241		
QY	3301	TAGGTTCCAGTATGAAAAATTCAAGCAGAACTACGTAGAAACAGAGGCCAAATTTGA	3360
Db	3301		
QY	3361	ATGCTATGCTTATGATTAGGGGTTTTGGCAACCTGAGGCTATTAACAAGTCTTCCGAA	3420
Db	3361		
QY	3421	GTAAATGTAAAGCATCTCTGAAATAAAAAACAGAAATATGAAAGTAGTTCAAGCTGTA	3480
Db	3421		
QY	3481	ATACAGATTTCTCTCATATCTGATTTTCAGATTAACAACGCTATGGGAAGTAGTC	3540
Db	3481		
QY	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGGTGAATTAAGG	3600
Db	3541		
QY	3601	AAGATAGTATGTTTCTGAAAAATGACATTAAGGAAAGTTCTGCTGTTTTTACAAAAGCG	3660
Db	3601		
QY	3661	TCCAGAAAGAGAGCTTAGACAGAGTCTTAGGCCCTTTCACCCATACATTTGGCTCAGG	3720
Db	3661		
QY	3721	GTACCGAAGAGGGCCCAAGAAATTAGAGTCTCGAAGAGAACTTATCTAGTGAGGATG	3780
Db	3721		
QY	3781	AAGAGCTTCCCTGCTTCCAAACCTGTTATTTGTAAAGTAAACATATACCTTCAGT	3840
Db	3781		
QY	3841	CTACTAGGCATAGACCGTTGCTACCGAGTGTCTTAAGAACAAGAGGAATTTAT	3900
Db	3841		
QY	3901	TATCATTTGAAGATTAACCTTAATGTAGTGCAGTAACAGGATATATTTGCAAGGATCTC	3960
Db	3901		
QY	3961	AGGAACATCACCTTATAGTGAAGAAACAAATGTTTGTAGTGTGTTTTCTTACAGTGCA	4020
Db	3961		
QY	4021	GTGAATTTGAACACTTTGACTGCAATTAACAACCCAGGATCTTTCTTGATTTGGTCTT	4080
Db	4021		
QY	4081	CCAAACAAATGAGGATCATGCTGTAACCCAGGGAGTGGTCTGAGTACAGGAATTTGG	4140
Db	4081		
QY	4141	TTTCAGATGATGAACAAAGAGAGAACGGGCTTGAAGAAATTAATCAAGAGCAAGCA	4200
Db	4141		
QY	4201	TGGATTTCAAACCTTAGGTGAGCAGCATCTGGGTGTGAGAGTGAACAAGCCTCTGAG	4260
Db	4201		
QY	4261	ACTGGTCAGGGCTATTCCTCTCGATGATGATTTTAACCACTCAGAGAGGATACCATGC	4320
Db	4261		
QY	4321	AAACATTAACCTGTATTAAGCTCCAGCAGAAATGGCTGAATAGAAAGCTGTGTAGAACGC	4380
Db	4321		
QY	4381	ATGGAGCCAGCCTTCTTAACAGCTACCTTCCATCATTAAGTACTCTTTCGCTTGAAG	4440
Db	4381		
QY	4441	ACCTGCGAAATCCAAACAAAGACATCAGAAAAAGAGTATTAATCTCAACAGAAAAATA	4500
Db	4441		
QY	4501	GTGAATTAACCTGTATTAAGCTCCAGAGGCTTTCGTGACAAAGTTGAGGTCTG	4560
Db	4501		
QY	4561	CAGATAGTTCTACAGTAAAAATTAAGAACCCAGAGTGGAAAGTCAATCCCTCTTAAT	4620
Db	4561		
QY	4621	GCCCATCATATGATATAGTGTGTCAATGACAGTGTGCTGGAGTCTTCAGAAATGA	4680
Db	4621		
QY	4681	ACTAACCATCTCAAGAGAGGCTCATTAAGTTGTGTGATGTGGAGAGCAACAGCTGGAAG	4740
Db	4681		
QY	4741	AGTCTGGGCGACAGATTTGACGGAACATTTACTTTGCCAAGCAAGATCTAGAGGAA	4800
Db	4741		
QY	4801	CCCCCTTACCCTGGAATCTGGAAATCAGCGCTCTCTGTGATGAAACCCGGAATCTGATCTTCTG	4860
Db	4801		
QY	4861	AAGACAGAGCCCCAGAGTCACTGCTGTGTGGCAATACATCTTCAACCTGCAATGA	4920
Db	4861		
QY	4921	AAGTTCCCAATTTGAAGTTGCAGAAATCTGCCAGAGTCCAGTGTCTCATATCACTG	4980
Db	4921		

Db		4921	AAGTTCCCAAFGTGAAGTTCGACAGATCTGGCCACAGATCCAGCTGCCTCATACTACTG	4986
OY		4981	ATACTGCTGGGTATATATCAATTGAGAAGAGTGAGCAGGAGAGACCAGAAITTGACAG	5040
Db		4981	ATACTGCTGGGTATATATCAATTGAGAAGAGTGAGCAGGAGAGACCAGAAITTGACAG	5040
OY		5041	CTTCAACGAAAGGCTCACAAAAGTAATGCCATGCTGCTGCTGCGCCCTGACCCAGAG	5100
Db		5041	CTTCAACGAAAGGCTCACAAAAGTAATGCCATGCTGCTGCTGCGCCCTGACCCAGAG	5100
OY		5101	AATTATATCTGTGACAGATTTGGCAGAAAAACACATCATCTAATCTATATTA	5160
Db		5101	AATTATATCTGTGACAGATTTGGCAGAAAAACACATCATCTAATCTATATTA	5160
OY		5161	CTGAAGACACTACTCATGTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGGACAC	5220
Db		5161	CTGAAGACACTACTCATGTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGGACAC	5220
OY		5221	TGAAATATTTTCTTAGGAATTCGGGAGGAGAAATGGGTAGTTAGCTATTTCTGGGTGACCC	5280
Db		5221	TGAAATATTTTCTTAGGAATTCGGGAGGAGAAATGGGTAGTTAGCTATTTCTGGGTGACCC	5280
OY		5281	AGCTATTTAAAGAAAGAAATATGCTGATGAGCATGATTTTGAATCTCAGAGAGATGTGG	5340
Db		5281	AGCTATTTAAAGAAAGAAATATGCTGATGAGCATGATTTTGAATCTCAGAGAGATGTGG	5340
OY		5341	TCATATGAGAAACACCCACCAAGGTCCAAAGCAGACNAGAAATCCCAGACAGAAAGATCT	5400
Db		5341	TCATATGAGAAACACCCACCAAGGTCCAAAGCAGACNAGAAATCCCAGACAGAAAGATCT	5400
OY		5401	TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTTCACCAATGCCCCACAGATCAACTGG	5460
Db		5401	TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTTCACCAATGCCCCACAGATCAACTGG	5460
OY		5461	AATGATGAGTACACTGTGTGCTGTCTGTGAGTGAAGAGCTTTTCATCATTCACCTCTG	5520
Db		5461	AATGATGAGTACACTGTGTGCTGTCTGTGAGTGAAGAGCTTTTCATCATTCACCTCTG	5520
OY		5521	GCAACAGGTGTCACCCCAATTTGTGTTGTCAGACAGATGCTTGACAGAGACAATGSGCT	5580
Db		5521	GCAACAGGTGTCACCCCAATTTGTGTTGTCAGACAGATGCTTGACAGAGACAATGSGCT	5580
OY		5581	TCATATGCAATTTGGCGCAGATGTGTGAGGCACCTGTGTGATACCCGAGAGTGGGTGGACA	5640
Db		5581	TCATATGCAATTTGGCGCAGATGTGTGAGGCACCTGTGTGATACCCGAGAGTGGGTGGACA	5640
OY		5641	GTGTAGACACTTACCAGATGTCGAGGAGAGCTGTGACACTTACTATATACCCAGATCCCCACA	5700
Db		5641	GTGTAGACACTTACCAGATGTCGAGGAGAGCTGTGACACTTACTATATACCCAGATCCCCACA	5700
OY		5701	GCCACTACTGA 5711	
Db		5701	GCCACTACTGA 5711	
RESULT 14				
LOCUS	181034		5914 bp	DNA linear PAT 10-JUN-1998
DEFINITION	Sequence 1 from patent US 5710001.			
VERSION	181034			
KEYWORDS	GI:3209324			
SOURCE ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 5914)			
AUTHORS	Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A., Hatsman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W. and Futreal,P.Andrew.			
TITLE	17q-linked breast and ovarian cancer susceptibility gene			
JOURNAL	Patent: US 5710001-A 1 20-JAN-1998;			
FEATURES	Location/Qualifiers			
source	I..5914			

BASE COUNT	2006 a	1156 c	1316 g	1436 t
ORIGIN	/organism="unknown"			
Query Match	99.9%; Score 5707.8; DB 6; Length 5914;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 5709;	Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	AGCTGCTGAGACCTTCTTGAGACCCCGCACAGGCTGTGGGGTTTCTCAGATAACTGGGCC	60	
DB	1	AGCTGCTGAGACCTTCTTGAGACCCCGCACAGGCTGTGGGGTTTCTCAGATAACTGGGCC	60	
QY	61	CCTGCGCTCAGAGAGCCCTTACACCTCTGCTGTGGSTAAAGTTCAATGACAGAAAGAA	120	
DB	61	CCTGCGCTCAGAGAGCCCTTACACCTCTGCTGTGGSTAAAGTTCAATGACAGAAAGAA	120	
QY	121	TGGAATTAATCGCTCTTGGCGCTTGAAGAAGTACAAAATGATTAATCTATGCGAATA	180	
DB	121	TGGAATTAATCGCTCTTGGCGCTTGAAGAAGTACAAAATGATTAATCTATGCGAATA	180	
QY	181	TCTTAGAGATGTCCTCATCTGTGTGAGATTGATCAAGAACTGTCTCCACAAAGTGTGACC	240	
DB	181	TCTTAGAGATGTCCTCATCTGTGTGAGATTGATCAAGAACTGTCTCCACAAAGTGTGACC	240	
QY	241	ACATATTTTGGCAAAATTTTGGCAGCTGAAACCTTCTCACACAGAAAGAGCCCTTCCACAGT	300	
DB	241	ACATATTTTGGCAAAATTTTGGCAGCTGAAACCTTCTCACACAGAAAGAGCCCTTCCACAGT	300	
QY	301	GTCCTTTATGTAACAAATGATTAATCAACAAAGAGGCTCAAGAAAGTACGAGATTACTC	360	
DB	301	GTCCTTTATGTAACAAATGATTAATCAACAAAGAGGCTCAAGAAAGTACGAGATTACTC	360	
QY	361	AACCTGTGTAAGAGCTATTGAAATCATTTTGTGCTTTACAGTTGACACAGATTGGAGT	420	
DB	361	AACCTGTGTAAGAGCTATTGAAATCATTTTGTGCTTTACAGTTGACACAGATTGGAGT	420	
QY	421	ATGCAAAACAGCTATTAATTTTCCAAAAGAAATAACTCTCTGACATCTTAAAGATG	480	
DB	421	ATGCAAAACAGCTATTAATTTTCCAAAAGAAATAACTCTCTGACATCTTAAAGATG	480	
QY	481	AAGTTTCATCATCCAAAGTATGGGCTACAAAACCGGCCAAAAGCTTCTACAGATG	540	
DB	481	AAGTTTCATCATCCAAAGTATGGGCTACAAAACCGGCCAAAAGCTTCTACAGATG	540	
QY	541	AACCCGAAATCCTCTTCCTTGAGAGAAACCACTCAGTGTCCAATCTCTAACTTGGAA	600	
DB	541	AACCCGAAATCCTCTTCCTTGAGAGAAACCACTCAGTGTCCAATCTCTAACTTGGAA	600	
QY	601	CTGTGAGAACTCTGAGAGCAAGAGCGGATACAACTCAAAAGACGTCTGTACATTTG	660	
DB	601	CTGTGAGAACTCTGAGAGCAAGAGCGGATACAACTCAAAAGACGTCTGTACATTTG	660	
QY	661	AATTGGGATCTGATTTCTTCTGAACATACCGTTAATAGGCAACTTATTGCACTGTGGAG	720	
DB	661	AATTGGGATCTGATTTCTTCTGAACATACCGTTAATAGGCAACTTATTGCACTGTGGAG	720	
QY	721	ATCAAGAATTTGTACAAATCACCCCTCAGAGAACAGGAGTGAATCAGTTTGGATTCTG	780	
DB	721	ATCAAGAATTTGTACAAATCACCCCTCAGAGAACAGGAGTGAATCAGTTTGGATTCTG	780	
QY	781	CAAAAAGAGCGCTTGTGAAATTTTCTGAGACGATGTACAAATTAATCTGAACATCATCAC	840	
DB	781	CAAAAAGAGCGCTTGTGAAATTTTCTGAGACGATGTACAAATTAATCTGAACATCATCAC	840	
QY	841	CCAGTAAATTAATGATTTGAAACCACTGAGAGGCTGTGACGCTGAGAGGCTATCCAAAAAGT	900	
DB	841	CCAGTAAATTAATGATTTGAAACCACTGAGAGGCTGTGACGCTGAGAGGCTATCCAAAAAGT	900	
QY	901	ATCAGGGTAGTCTGTTTCAAACTTGATGTGAGACCATGTGGACAAATCTCATGCCA	960	
DB	901	ATCAGGGTAGTCTGTTTCAAACTTGATGTGAGACCATGTGGACAAATCTCATGCCA	960	
QY	961	GCTCATTTACAGCATAGAACAGCAGTTTATTACTACTAAAGACAGATGAATGTAGAAA	1020	

|||||  
Db 961 GCTCATTCACAGCATGAGAACAGCAGTTATTACTACATAAGACAGATGAATGATGAAA 1020  
QY 1021 AGGCGAATTCCTGTAATATAAAGCAAAACAGCCTGGCTTAGCAAGAGCCAACTAATACAT 1080  
Db 1021 AGGCGAATTCCTGTAATATAAAGCAAAACAGCCTGGCTTAGCAAGAGCCAACTAATACAT 1080  
QY 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACTCCAGCAGACAGAAAAAAGTAG 1140  
Db 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACTCCAGCAGACAGAAAAAAGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGANTGGAATAGCAGAACTGCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGANTGGAATAGCAGAACTGCCATGCT 1200  
QY 1201 CAGAGATCCTAGAGTACTGAAAGATGTTCTTGATTAACACTAAATGAGCAATTCGA 1260  
Db 1201 CAGAGATCCTAGAGTACTGAAAGATGTTCTTGATTAACACTAAATGAGCAATTCGA 1260  
QY 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTATGAACGTGTTAGTTCTGATGACTCAGATGATG 1320  
Db 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTATGAACGTGTTAGTTCTGATGACTCAGATGATG 1320  
QY 1321 GGGAGTGTGATCAATGCCAAGTACGTGATGATTGGACGTTCTAATAGTAGTAGTG 1380  
Db 1321 GGGAGTGTGATCAATGCCAAGTACGTGATGATTGGACGTTCTAATAGTAGTAGTG 1380  
QY 1381 AATATTCTGTTCTTCAGAGAAATAGACTTACGCGCAGTACGTCATGAGGCTTTAA 1440  
Db 1381 AATATTCTGTTCTTCAGAGAAATAGACTTACGCGCAGTACGTCATGAGGCTTTAA 1440  
QY 1441 TATGTAAAGTGAAGAAGTTCACCTCAATTCAGTAGAGAGTAATATTGAACACAAATAT 1500  
Db 1441 TATGTAAAGTGAAGAAGTTCACCTCAATTCAGTAGAGAGTAATATTGAACACAAATAT 1500  
QY 1501 TTGGGAAAACCTATCGGAAAGGCAAGCCTCCCAACTTAAGCCATGTAAGTAAATC 1560  
Db 1501 TTGGGAAAACCTATCGGAAAGGCAAGCCTCCCAACTTAAGCCATGTAAGTAAATC 1560  
QY 1561 TAAATTATGAGAGCATTTGTTACTGAGGCACAGATTAATACAAAGCGTCCCTCAANA 1620  
Db 1561 TAAATTATGAGAGCATTTGTTACTGAGGCACAGATTAATACAAAGCGTCCCTCAANA 1620  
QY 1621 AATTTAAAGCGTAAAGAGGAGACCTACATCAGGCCCTTACCTCGAGAGATTTATCAAGAA 1680  
Db 1621 AATTTAAAGCGTAAAGAGGAGACCTACATCAGGCCCTTACCTCGAGAGATTTATCAAGAA 1680  
QY 1681 CAGATTTGGCAGTTCAAAAAGACCTCTGAAATGATAAATCAGGGAACCTAACCAAGCGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAAGACCTCTGAAATGATAAATCAGGGAACCTAACCAAGCGAGC 1740  
QY 1741 AGAATGGTCAAGTGAATATTACTAATAGTGGTCAATGAGAAATTAACAAAGAGTAT 1800  
Db 1741 AGAATGGTCAAGTGAATATTACTAATAGTGGTCAATGAGAAATTAACAAAGAGTAT 1800  
QY 1801 CTATTCAGAAATGAGAAAAATCTTAACCAATAGAACTCGAAAAAGAAATCTGCTTCA 1860  
Db 1801 CTATTCAGAAATGAGAAAAATCTTAACCAATAGAACTCGAAAAAGAAATCTGCTTCA 1860  
QY 1861 AAACGAAAGCTGAACCTATTAAGCAGAGTAAAGCAATATGAACTGAAATTAATATCC 1920  
Db 1861 AAACGAAAGCTGAACCTATTAAGCAGAGTAAAGCAATATGAACTGAAATTAATATCC 1920  
QY 1921 ACAATTCAAAAGACCTAAAGAAATAGGCTGAGAGAGAGTCTTCAACAGGATATTC 1980  
Db 1921 ACAATTCAAAAGACCTAAAGAAATAGGCTGAGAGAGAGTCTTCAACAGGATATTC 1980  
QY 1981 ATGCGCTTGAACCTAGTACAGTAAATCTAAGCCACCTAATTTGACTGAATTCGAA 2040  
Db 1981 ATGCGCTTGAACCTAGTACAGTAAATCTAAGCCACCTAATTTGACTGAATTCGAA 2040  
QY 2041 TTGATAGTGTGTAGCAGTGAAGAGATTAAGAAAAAGTACACCAATGCGCAGTCA 2100  
Db 2041 TTGATAGTGTGTGTAGCAGTGAAGAGATTAAGAAAAAGTACACCAATGCGCAGTCA 2100

Db 2041 TTGATAGTGTGTGTAGCAGTGAAGAGATTAAGAAAAAGTACACCAATGCGCAGTCA 2100  
QY 2101 GGCACAGAGAAACCTCAACTCATGTAAGTAAAGAAAGAACTGCAACTGGAGCCAAAGAA 2160  
Db 2101 GGCACAGAGAAACCTCAACTCATGTAAGTAAAGAAAGAACTGCAACTGGAGCCAAAGAA 2160  
QY 2161 GTAAACAGCCAAATGAAGACAGCAAGTAAAGACATGACATGATCTTCCAGACTGA 2220  
Db 2161 GTAAACAGCCAAATGAAGACAGCAAGTAAAGACATGACATGATCTTCCAGACTGA 2220  
QY 2221 AGTTAACAATGACACTGTTCTTTACTAAGTGTTCAAATACCAAGTAAATTAAGAT 2280  
Db 2221 AGTTAACAATGACACTGTTCTTTACTAAGTGTTCAAATACCAAGTAAATTAAGAT 2280  
QY 2281 TTGTCATCTGAGCTTCCAGAGAAAGAAAGAAAGAAAGTAAACAGTTAAAGTCT 2340  
Db 2281 TTGTCATCTGAGCTTCCAGAGAAAGAAAGAAAGTAAACAGTTAAAGTCT 2340  
QY 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAAGT 2400  
Db 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAAGT 2400  
QY 2401 AAAGATCTGTAGAGAGTACAGATATTTCAATGCTGCTGACTGATTAATGAGCTCAG 2460  
Db 2401 AAAGATCTGTAGAGAGTACAGATATTTCAATGCTGCTGACTGATTAATGAGCTCAG 2460  
QY 2461 AAAGATCTGTAGAGAGTACAGATATTTCAATGCTGCTGACTGATTAATGAGCTCAG 2520  
Db 2461 AAAGATCTGTAGAGAGTACAGATATTTCAATGCTGCTGACTGATTAATGAGCTCAG 2520  
QY 2521 GTGTGAGTCAAGTGTGACAGATTTGAAACCCCAAGGACTAATTCATGCTGTTCCAAAG 2580  
Db 2521 GTGTGAGTCAAGTGTGACAGATTTGAAACCCCAAGGACTAATTCATGCTGTTCCAAAG 2580  
QY 2581 ATTAATGAATGACACAGAAAGGCTTTAGATATCATTTGGACATGAAATTAACACAGTC 2640  
Db 2581 ATTAATGAATGACACAGAAAGGCTTTAGATATCATTTGGACATGAAATTAACACAGTC 2640  
QY 2641 GGGAAACAGCATTTGAAGTGAAGAAAGTCACTGTTGCTGATTTGGCAAGATACAT 2700  
Db 2641 GGGAAACAGCATTTGAAGTGAAGAAAGTCACTGTTGCTGATTTGGCAAGATACAT 2700  
QY 2701 TCAAGGTTTCAAAAGCCGCACTATTTGCTGTTTCAATTCAGAGAAATGCGAAGAG 2760  
Db 2701 TCAAGGTTTCAAAAGCCGCACTATTTGCTGTTTCAATTCAGAGAAATGCGAAGAG 2760  
QY 2761 AATGTGCAACATTTCTGCCCCACTCTGCGTCTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
Db 2761 AATGTGCAACATTTCTGCCCCACTCTGCGTCTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
QY 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGATGATGATATCAAGCCGTATC 2880  
Db 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGATGATGATATCAAGCCGTATC 2880  
QY 2881 AGACAGTTAAATACACTGAGGCTTTCTGTTGTTGCTGAGAAAGTAAAGCCAGTTGATA 2940  
Db 2881 AGACAGTTAAATACACTGAGGCTTTCTGTTGTTGCTGAGAAAGTAAAGCCAGTTGATA 2940  
QY 2941 ATGCCAAATGTAGTATCAAGAGAGCTTAGGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
Db 2941 ATGCCAAATGTAGTATCAAGAGAGCTTAGGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
QY 3001 ACGAACCTGAGCTATTTACCCCAATTAACATGCACTTTTACAAAACCATATGCTATAC 3060  
Db 3001 ACGAACCTGAGCTATTTACCCCAATTAACATGCACTTTTACAAAACCATATGCTATAC 3060  
QY 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAATAATCGCTAGAG 3120  
Db 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAATAATCGCTAGAG 3120  
QY 3121 AAAACCTTTGAGAGCAATTCATGTCACCTGAAAGAGAAATGGAATGGAACATTTCCA 3180  
Db 3121 AAAACCTTTGAGAGCAATTCATGTCACCTGAAAGAGAAATGGAATGGAACATTTCCA 3180

OY	3181	GTACAGTGAACAATTAGCCGTATAAATAGAGAAAAAGTTTTAAAGAACCCAGCT	3240
Db	3181	GTACAGTGAACAATTAGCCGTATAAATAGAGAAAAAGTTTTAAAGAACCCAGCT	3240
OY	3241	CAAGCAATTAATTAAGAACTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTTAATATAA	3300
Db	3241	CAAGCAATTAATTAAGAACTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTTAATATAA	3300
OY	3301	TAGGTTCCAGTGAAGAAACATTCACAGACAATAGTAGAAACAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTGAAGAAACATTCACAGACAATAGTAGAAACAGAGGGCCAAATTTGA	3360
OY	3361	ATGCTATGCTTAGATTAAGGGTTTTTGCACTGAGGCTATAAACAAAGTCTTCGGA	3420
Db	3361	ATGCTATGCTTAGATTAAGGGTTTTTGCACTGAGGCTATAAACAAAGTCTTCGGA	3420
OY	3421	GTAATTTGAACATCTGCAATTAATAAAAAAGCAGATATTGAAGAAGTGGTTCAGCTGTA	3480
Db	3421	GTAATTTGAACATCTGCAATTAATAAAAAAGCAGATATTGAAGAAGTGGTTCAGCTGTA	3480
OY	3481	ATACAGATTTCTCOCATATCTGATTTTCAGATTAATTAGAACACAGCTATGGGAATGTC	3540
Db	3481	ATACAGATTTCTCOCATATCTGATTTTCAGATTAATTAGAACACAGCTATGGGAATGTC	3540
OY	3541	ATGCATCTCAGGTTGTTGTGAGACACCTGATGACCTGTTAGATGATGGTAATAAAG	3600
Db	3541	ATGCATCTCAGGTTGTTGTGAGACACCTGATGACCTGTTAGATGATGGTAATAAAG	3600
OY	3601	AAGATTAAGTTTGTCTGAATAATGACATTTAAGAAACCTTCGCTTTTATAGCAAAAGCG	3660
Db	3601	AAGATTAAGTTTGTCTGAATAATGACATTTAAGAAACCTTCGCTTTTATAGCAAAAGCG	3660
OY	3661	TTCCAGAAAGAGACCTTAGCAGGAGTCTTAGCCCTTACCCATACCATTTGGCTAGG	3720
Db	3661	TTCCAGAAAGAGACCTTAGCAGGAGTCTTAGCCCTTACCCATACCATTTGGCTAGG	3720
OY	3721	GTTACCGGAAGGGGGCCACAAGAAATTAAGATCTCAGAGAAGAACTTATCTAGTAGATG	3780
Db	3721	GTTACCGGAAGGGGGCCACAAGAAATTAAGATCTCAGAGAAGAACTTATCTAGTAGATG	3780
OY	3781	AAGAGCTTCCCTGCTTCCACACCTGTTATTTGGTAAGTAACATATACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTTCCACACCTGTTATTTGGTAAGTAACATATACCTTCTCAGT	3840
OY	3841	CTACTAGGCATAGACCGTGTCTACGAGGTGCTGTAGAAACACAGAGAGAAATTTAT	3900
Db	3841	CTACTAGGCATAGACCGTGTCTACGAGGTGCTGTAGAAACACAGAGAGAAATTTAT	3900
OY	3901	TATCATTTGAAGAAATAGCTTAATATACTCAGATAACACAGTAATATTGGCAAGGCAATCTC	3960
Db	3901	TATCATTTGAAGAAATAGCTTAATATACTCAGATAACACAGTAATATTGGCAAGGCAATCTC	3960
OY	3961	AGGAACATTCACCTTAGTAGAGAAACAAATGCTGCTAGCTGTTTTCTTCACAGTGA	4020
Db	3961	AGGAACATTCACCTTAGTAGAGAAACAAATGCTGCTAGCTGTTTTCTTCACAGTGA	4020
OY	4021	GTAATTTGAAGAACTTACTGCAAAATCAAAACACCCAGATCTCTTCTTATTTGGTCTCT	4080
Db	4021	GTAATTTGAAGAACTTACTGCAAAATCAAAACACCCAGATCTCTTCTTATTTGGTCTCT	4080
OY	4081	CCAAACAATTAAGGCATACGTCTGAAAAGCCAGGGAGTTGGCTGAGTGAACAAAGAAATTTGG	4140
Db	4081	CCAAACAATTAAGGCATACGTCTGAAAAGCCAGGGAGTTGGCTGAGTGAACAAAGAAATTTGG	4140
OY	4141	TTTTCAGATGATGAAGAAAGAGAAACGGGCTTTGGAAAGAAATTAATCAAGAAAGACAAAGCA	4200
Db	4141	TTTTCAGATGATGAAGAAAGAGAAACGGGCTTTGGAAAGAAATTAATCAAGAAAGACAAAGCA	4200
OY	4201	TGGATTCAAACTTAGTGTGAAGACAGCATCTGGGTGTGAGTGAAGAAACAGCGTCTCTGAAG	4260
Db	4201	TGGATTCAAACTTAGTGTGAAGACAGCATCTGGGTGTGAGTGAAGAAACAGCGTCTCTGAAG	4260

[illegible]

|||||  
Db 5341 TCATGAGAAACCAACAGAGTCACAAAGGAGACAGAGATCCACAGACAGAAAGATCT 5400  
5401 TCAGGGGGGCTAGAAATCTGTGTCTATGAGGCCCTTACACCAATGCCACAGATCAACTGG 5460  
5402 TCAGGGGGGCTAGAAATCTGTGTCTATGAGGCCCTTACACCAATGCCACAGATCAACTGG 5460  
5461 AATGATGATGACAGCTGTGTGTCTCTGTGTGAAGAGCTTTCATCTACCTG 5520  
5462 AATGATGATGACAGCTGTGTGTCTCTGTGTGAAGAGCTTTCATCTACCTG 5520  
5463 AATGATGATGACAGCTGTGTGTCTCTGTGTGAAGAGCTTTCATCTACCTG 5520  
5521 GCACAGGTGTCCACCAATTTGTGTGTGACAGCAGATGCTTGACAGAGAGCAATGGCT 5580  
5522 GCACAGGTGTCCACCAATTTGTGTGTGACAGCAGATGCTTGACAGAGAGCAATGGCT 5580  
5581 TCCATGCAATTTGGGAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5640  
5582 TCCATGCAATTTGGGAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5640  
5583 TCCATGCAATTTGGGAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5640  
5584 TCCATGCAATTTGGGAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5640  
5641 GTGTAGCAGCTCTACAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5700  
5642 GTGTAGCAGCTCTACAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5700  
5643 GTGTAGCAGCTCTACAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5700  
5644 GTGTAGCAGCTCTACAGAGTGTGTGAGGAGCAGCTGTGTGAGAGAGAGAGAGAGAGAG 5700  
Qy 5701 GCCACTACTGA 5711  
Db 5701 GCCACTACTGA 5711

RESULT 15  
AR048660  
LOCUS AR048660 5711 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 4 from patent US 5821328.  
ACCESSION AR048660  
VERSION AR048660.1 GI:59711003  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 5711)  
AUTHORS King, M.-C., Friedman, L., Ostermeyer, B., Rowell, S., Lynch, E., Szabo, C. and Lee, M.  
TITLE Genetic markers for breast, ovarian, and prostatic cancer  
JOURNAL Patent: US 5821328-A 4 13-Oct-1998;  
FEATURES  
source 1. 5711  
location/Qualifiers  
BASE COUNT 1956 a 1099 c 1275 g 1381 t  
ORIGIN

Query Match 99.9%; Score 5706.2; DB 6; Length 5711;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 5708; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 301 GTCTTATATGTAGAAATGATATAACCAAAAGAGCCTACAGAAAGTACAGATTTAGTC 360  
Db 301 GTCTTATATGTAGAAATGATATAACCAAAAGAGCCTACAGAAAGTACAGATTTAGTC 360  
Qy 361 AACTTGTGAAGAGCTATTGAAATCAATTGTGCTTTTCAGCTTGGACAGAGTTTGGAGT 420  
Db 361 AACTTGTGAAGAGCTATTGAAATCAATTGTGCTTTTCAGCTTGGACAGAGTTTGGAGT 420  
Qy 421 ATGCAACAGCTATTAATTTTGCACAAAAGAAATAAATCTCTGTAACATCTTAAAGATG 480  
Db 421 ATGCAACAGCTATTAATTTTGCACAAAAGAAATAAATCTCTGTAACATCTTAAAGATG 480  
Qy 481 AAGTTTCTATTCATCCAAAGATATGAGGCTACAGAAACCGTGCCAAAGACTTTACAGAGTG 540  
Db 481 AAGTTTCTATTCATCCAAAGATATGAGGCTACAGAAACCGTGCCAAAGACTTTACAGAGTG 540  
Qy 541 AACCCGAAATCCCTCTCTTTCAGAGAAACAGTGTCAAGTGTCCAACTCTTAACCTTGGA 600  
Db 541 AACCCGAAATCCCTCTCTTTCAGAGAAACAGTGTCAAGTGTCCAACTCTTAACCTTGGA 600  
Qy 601 CTGTGAGAGCTGTGAGAGCAAGAGGAGATACACCTCAAAAGAGCTGTGTACATTTG 660  
Db 601 CTGTGAGAGCTGTGAGAGCAAGAGGAGATACACCTCAAAAGAGCTGTGTACATTTG 660  
Qy 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGCACTTAATGCAAGTGTGGAG 720  
Db 661 AATTGGATCTGATTTCTTGAAGATACCGTTAATAAGCACTTAATGCAAGTGTGGAG 720  
Qy 721 ATCAAGAAATTTTCAAAATACCCCTCAAGAGACAGAGATGAATATCAGTTTGGATTCTG 780  
Db 721 ATCAAGAAATTTTCAAAATACCCCTCAAGAGACAGAGATGAATATCAGTTTGGATTCTG 780  
Qy 781 CAAAAGAGGCGCTTGTGAATTTTCTGAGAGCGAGTGAACAAATPACTGAAATCATCAAC 840  
Db 781 CAAAAGAGGCGCTTGTGAATTTTCTGAGAGCGAGTGAACAAATPACTGAAATCATCAAC 840  
Qy 841 CCAGTAATTAATGATTTGACACACACAGAGGCTGAGAGGATCCAGAAAGT 900  
Db 841 CCAGTAATTAATGATTTGACACACACAGAGGCTGAGAGGATCCAGAAAGT 900  
Qy 901 ATCAGGATGATGCTTTTCAAACTTGCATGTGAGAGGCTGTGGCAAAATATCTCATGCA 960  
Db 901 ATCAGGATGATGCTTTTCAAACTTGCATGTGAGAGGCTGTGGCAAAATATCTCATGCA 960  
Qy 961 GCTCATTAACAGATGAGAGACAGAGCTTATTTACTCAATAAGACAGAAATGATGTGAA 1020  
Db 961 GCTCATTAACAGATGAGAGACAGAGCTTATTTACTCAATAAGACAGAAATGATGTGAA 1020  
Qy 1021 AGGCTGAATTCCTGTAATTAAGCAACAGGCTGTAGCAGAGGACCAATACAGAT 1080  
Db 1021 AGGCTGAATTCCTGTAATTAAGCAACAGGCTGTAGCAGAGGACCAATACAGAT 1080  
Qy 1081 GGGCTGGAAGTAAGCAACATGTAATGATAGGGGAGCTCCACACAGAAAAAGGTAG 1140  
Db 1081 GGGCTGGAAGTAAGCAACATGTAATGATAGGGGAGCTCCACACAGAAAAAGGTAG 1140  
Qy 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATAGCAGAAATCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATAGCAGAAATCCATGCT 1200  
Qy 1201 CAGAGAACTCTAGAGATCTGAAGATGCTTGTGATTAACACTAAATAGCAGATTCAGA 1260  
Db 1201 CAGAGAACTCTAGAGATCTGAAGATGCTTGTGATTAACACTAAATAGCAGATTCAGA 1260  
Qy 1261 AAGTTAATGATGCTTTTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
Db 1261 AAGTTAATGATGCTTTTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
Qy 1321 GGGAGCTGATCAAAATGCAAGAGTATGATGATGATGATGATGATGATGATGATGATG 1380  
Db 1321 GGGAGCTGATCAAAATGCAAGAGTATGATGATGATGATGATGATGATGATGATGATG 1380  
Qy 1381 AATATTTCTGCTTCTGAGAGAAATAGACTTACTGGCCAGTGTACTCTCATGAGGCTTTAA 1440



[illegible]

Db	2461	AAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAGGCAAAAACACCAATAAT	2520
QY	2521	GTGTGAGTCAAGTGTGACAGACTTTGGAAACCCCAAGGGAGCTAATTCATGCTGTTCCAAAG	2580
Db	2521	GTGTGAGTCAAGTGTGACAGACTTTGGAAACCCCAAGGGAGCTAATTCATGCTGTTCCAAAG	2580
QY	2551	ATATATAGAAATGACACAGAGAGCTTTAAAGTATCCATTGTGGACATGAACTTAAACACAGTC	2640
Db	2551	ATATATAGAAATGACACAGAGAGCTTTAAAGTATCCATTGTGGACATGAACTTAAACACAGTC	2640
QY	2641	GGGAACACAGATATGAAATGGAAGAAAGTAACTGATGCTCAGTATTTTGCAGAAATCAT	2700
Db	2641	GGGAACACAGATATGAAATGGAAGAAAGTAACTGATGCTCAGTATTTTGCAGAAATCAT	2700
QY	2701	TCAGAGTTTCAAAGGGCCAGTCAATTTGCTCTGTTTTCAATCCAGAAATGCAGAAAGAG	2760
Db	2701	TCAGAGTTTCAAAGGGCCAGTCAATTTGCTCTGTTTTCAATCCAGAAATGCAGAAAGAG	2760
QY	2761	AATGTGCAACATTTCTGTGCCCACTCTGGGTCTTAAGAAACAAAGTCCAAAGTCATT	2820
Db	2761	AATGTGCAACATTTCTGTGCCCACTCTGGGTCTTAAGAAACAAAGTCCAAAGTCACTT	2820
QY	2821	TTGAATGTGAACAAAGGAAGAAATCAAGGAAGAATGACTCTAATATCAAGCTGTAC	2880
Db	2821	TTGAATGTGAACAAAGGAAGAAATCAAGGAAGAATGACTCTAATATCAAGCTGTAC	2880
QY	2881	AGACAGTTAATATCACTCAGAGCTTTCTGTGTGGTGCACAAAGATTAAGCAGCTTGATA	2940
Db	2881	AGACAGTTAATATCACTCAGAGCTTTCTGTGTGGTGCACAAAGATTAAGCAGCTTGATA	2940
QY	2941	ATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCCAGAGCA	3000
Db	2941	ATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCCAGAGCA	3000
QY	3001	ACGAAACCTGGAGCTATTTCTCCAAATTAACATGAGACTTTTCAAAACCCATATGTGTATAC	3060
Db	3001	ACGAAACCTGGAGCTATTTCTCCAAATTAACATGAGACTTTTCAAAACCCATATGTGTATAC	3060
QY	3061	CACACACTTTTCCCATCAAGTCAATTTGTTTAAACCTAATGTGAAGAAAAATGTGCTAGAGG	3120
Db	3061	CACACACTTTTCCCATCAAGTCAATTTGTTTAAACCTAATGTGAAGAAAAATGTGCTAGAGG	3120
QY	3121	AAAACTTTGAGGAACATTCAATGTGCACCTGAAAGAGAATGGAATGTAGAACTTCCAA	3180
Db	3121	AAAACTTTGAGGAACATTCAATGTGCACCTGAAAGAGAATGGAATGTAGAACTTCCAA	3180
QY	3181	GTACAGTGTAGACAACTTATGCCGTATTAATCACTTAAGAGAAATGTTTTTAAGAGCCAGCT	3240
Db	3181	GTACAGTGTAGACAACTTATGCCGTATTAATCACTTAAGAGAAATGTTTTTAAGAGCCAGCT	3240
QY	3241	CAGCAATATTAATGAAGTAGTGTCCAGTACTAATGAAAGTGGGTCCAGTTTAAATGAAA	3300
Db	3241	CAGCAATATTAATGAAGTAGTGTCCAGTACTAATGAAAGTGGGTCCAGTTTAAATGAAA	3300
QY	3301	TAGGTTCCAGTATGAGAAACACTTCAAGAGAACTAGTAGAAGAAACAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTATGAGAAACACTTCAAGAGAACTAGTAGAAGAAACAGAGGGCCAAATTTGA	3360
QY	3361	ATGCTATGCTTGTAGATTAGGGGTTTTGCAACCTGAGAGCTATTAACCAAAAGTCTTCTGGAA	3420
Db	3361	ATGCTATGCTTGTAGATTAGGGGTTTTGCAACCTGAGAGCTATTAACCAAAAGTCTTCTGGAA	3420
QY	3421	GTAATTTGAAGCAATCCGTAATAAAAAAGCAAGAAATGTGAAGAGTGTGACAGCTTTA	3480
Db	3421	GTAATTTGAAGCAATCCGTAATAAAAAAGCAAGAAATGTGAAGAGTGTGACAGCTTTA	3480
QY	3481	ATACAGATTTCTCCAAATCTGATTCAGATTACTTAGAACACAGCTATGGAAGTGTGTC	3540
Db	3481	ATACAGATTTCTCCAAATCTGATTCAGATTACTTAGAACACAGCTATGGAAGTGTGTC	3540
QY	3541	ATGCATCTCAAGTTTGTCTGTGAGACACCTGATGACCTGTTAGATGATGTGAAATTAAGG	3600
Db	3541	ATGCATCTCAAGTTTGTCTGTGAGACACCTGATGACCTGTTAGATGATGTGAAATTAAGG	3600

QY 3601 AAGATAGTATTTGCTGAAAATGACATTAAGAAAGTTCTGCTTTTATGCAAAAGCG 3660  
DB 3601 AAGATAGTATTTGCTGAAAATGACATTAAGAAAGTTCTGCTTTTATGCAAAAGCG 3660  
QY 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAGG 3720  
DB 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAGG 3720  
QY 3721 GTTACGAGAGAGGGGSCAAGAAATTAGAGTCTCTGAGAGAGACTATCTAGTGGAGAG 3780  
DB 3721 GTTACGAGAGAGGGGSCAAGAAATTAGAGTCTCTGAGAGAGACTATCTAGTGGAGAG 3780  
QY 3781 AAGAGCTTCCCTGCTTCCACACTTGTATTTGGTAAGTAACAAATATACCTTCTCAGT 3840  
DB 3781 AAGAGCTTCCCTGCTTCCACACTTGTATTTGGTAAGTAACAAATATACCTTCTCAGT 3840  
QY 3841 CTACTAGGCATAGCAGCCGTTCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTTAT 3900  
DB 3841 CTACTAGGCATAGCAGCCGTTCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTTAT 3900  
QY 3901 TATCATTAAGAAATAGCTTAATAGACTGACATACCCAGGTAAATTTGGCAAAAGGATCTC 3960  
DB 3901 TATCATTAAGAAATAGCTTAATAGACTGACATACCCAGGTAAATTTGGCAAAAGGATCTC 3960  
QY 3961 AGGAACTCACTTATGAGGAAACAAAATGTTCTGCTAGCTTGTCTTCTCAGAGTCA 4020  
DB 3961 AGGAACTCACTTATGAGGAAACAAAATGTTCTGCTAGCTTGTCTTCTCAGAGTCA 4020  
QY 4021 GTGATTTGGAAGACTTGTACTGCAAAATACAAACACCAGAGATCTTCTTGATGGTCTT 4080  
DB 4021 GTGATTTGGAAGACTTGTACTGCAAAATACAAACACCAGAGATCTTCTTGATGGTCTT 4080  
QY 4081 CCAAAACAATGAGGCATCAGTCTGAAAGCCAGGGAGTGTCTGAGTACAGAGAAATGG 4140  
DB 4081 CCAAAACAATGAGGCATCAGTCTGAAAGCCAGGGAGTGTCTGAGTACAGAGAAATGG 4140  
QY 4141 TTTTCAGATGATGAAGAAAGAGAAAGGGCTTGGAAAGAAATATTCAGAGAGCA 4200  
DB 4141 TTTTCAGATGATGAAGAAAGAGAAAGGGCTTGGAAAGAAATATTCAGAGAGCA 4200  
QY 4201 TGGATTTAAACTTATGATGAGACACATCTGGGTGTGAGAGTGAACAAGCTCTGAG 4260  
DB 4201 TGGATTTAAACTTATGATGAGACACATCTGGGTGTGAGAGTGAACAAGCTCTGAG 4260  
QY 4261 ACTGCTCAGGGCTATCTCTCAGAGTGCATTTTAACCACTCAGCAGAGGGATACCATGC 4320  
DB 4261 ACTGCTCAGGGCTATCTCTCAGAGTGCATTTTAACCACTCAGCAGAGGGATACCATGC 4320  
QY 4321 AACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAACCTAGAAAGCTGTGAACAGC 4380  
DB 4321 AACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAACCTAGAAAGCTGTGAACAGC 4380  
QY 4381 ATGGAGGCCAGCCCTTTCACAGCTACCTTCCATCATATAGTACTCTTCTGAGG 4440  
DB 4381 ATGGAGGCCAGCCCTTTCACAGCTACCTTCCATCATATAGTACTCTTCTGAGG 4440  
QY 4441 ACCGTGCAAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTCAGAGAAAGTA 4500  
DB 4441 ACCGTGCAAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTCAGAGAAAGTA 4500  
QY 4501 GTGATTAACCTTATTAAGCCAGAAATCCAGAAAGGCTTCTGTGACAAAGTTGAGGTGTG 4560  
DB 4501 GTGATTAACCTTATTAAGCCAGAAATCCAGAAAGGCTTCTGTGACAAAGTTGAGGTGTG 4560  
QY 4561 CAGATTAATCTTACAGTAAAAATTAAGAACAGAGTGGAAAGTGCATCCCTCTTAAT 4620  
DB 4561 CAGATTAATCTTACAGTAAAAATTAAGAACAGAGTGGAAAGTGCATCCCTCTTAAT 4620  
QY 4621 GCCCATATTAAGATAGGTGATACATGCACAGTGTCTGGAGTCTTCAGAAATAGAA 4680  
DB 4621 GCCCATATTAAGATAGGTGATACATGCACAGTGTCTGGAGTCTTCAGAAATAGAA 4680

QY 4681 ACTACCATCTTCAGAGAGGCTCATTAAGTTGTTGATGTGAGAGAGCAACAGCTGGAAG 4740  
DB 4681 ACTACCATCTTCAGAGAGGCTCATTAAGTTGTTGATGTGAGAGAGCAACAGCTGGAAG 4740  
QY 4741 AGCTGGGCGCACAGATTTTACGGAAGATCTTACTTGGCCAGGCAAGATCTAGAGGAA 4800  
DB 4741 AGCTGGGCGCACAGATTTTACGGAAGATCTTACTTGGCCAGGCAAGATCTAGAGGAA 4800  
QY 4801 CCCCCTTACTGGAATCTGGAATCCAGCCCTTCTGAGTACGACCCGAATCTGATCTCTG 4860  
DB 4801 CCCCCTTACTGGAATCTGGAATCCAGCCCTTCTGAGTACGACCCGAATCTGATCTCTG 4860  
QY 4861 AAGACAGAGCCCGCAGAGTCAAGTCTGTTGGCAACATACCAATCTTCAACCTCTGCATTTGA 4920  
DB 4861 AAGACAGAGCCCGCAGAGTCAAGTCTGTTGGCAACATACCAATCTTCAACCTCTGCATTTGA 4920  
QY 4921 AAGTTCCCAATTTGAABTTCAGAAATCTGCCAGAGTCCAGCTGTCTCATTAATCTG 4980  
DB 4921 AAGTTCCCAATTTGAABTTCAGAAATCTGCCAGAGTCCAGCTGTCTCATTAATCTG 4980  
QY 4981 ATACTGCTGGGATTAATGCAATGGAAGAGTGTGAGAGAGGAGAGCCAGAAATTCAG 5040  
DB 4981 ATACTGCTGGGATTAATGCAATGGAAGAGTGTGAGAGAGGAGAGCCAGAAATTCAG 5040  
QY 5041 CTTCAACAGAAAGGCTCAACAAAAGAAATGTCATGATGTTGCTGAGCTGACCCAGAG 5100  
DB 5041 CTTCAACAGAAAGGCTCAACAAAAGAAATGTCATGATGTTGCTGAGCTGACCCAGAG 5100  
QY 5101 AATTATCTGCTGTGACAGATTTTGGCCAGAAACACATCTTAACTTAATTA 5160  
DB 5101 AATTATCTGCTGTGACAGATTTTGGCCAGAAACACATCTTAACTTAATTA 5160  
QY 5161 CTGAAGAGACTACTCATGTTGTTGAAGAACAGATGCTGAGTGTGTGTAACAGGAC 5220  
DB 5161 CTGAAGAGACTACTCATGTTGTTGAAGAACAGATGCTGAGTGTGTGTAACAGGAC 5220  
QY 5221 TGAATATTTTCTAGGAATTTGCGGAGAGAAATGGTAGTATTTCTGGGTGACCC 5280  
DB 5221 TGAATATTTTCTAGGAATTTGCGGAGAGAAATGGTAGTATTTCTGGGTGACCC 5280  
QY 5281 AGCTATTAAGAAAGAAATGCTGAATGAGATGATTTTGAAGTCAAGAGAGATGTG 5340  
DB 5281 AGCTATTAAGAAAGAAATGCTGAATGAGATGATTTTGAAGTCAAGAGAGATGTG 5340  
QY 5341 TCAATGGAAGAAACACCAAGGTCGAAAGCAGAGAGAGAAATCCAGAGAGAAATCT 5400  
DB 5341 TCAATGGAAGAAACACCAAGGTCGAAAGCAGAGAGAGAAATCCAGAGAGAAATCT 5400  
QY 5401 TCAGGGGCTAGAAATCTGTTGCTATGAGGCTTCCACCAATGCTCCAGATCAACTG 5460  
DB 5401 TCAGGGGCTAGAAATCTGTTGCTATGAGGCTTCCACCAATGCTCCAGATCAACTG 5460  
QY 5461 AATGAGTGTACACTGTGTGCTCTGTGGTGAAGAGCTTCTCATTAACCTG 5520  
DB 5461 AATGAGTGTACACTGTGTGCTCTGTGGTGAAGAGCTTCTCATTAACCTG 5520  
QY 5521 GCACAGGTGTCACCCCAATTTGTTGTGAGCAGCAGATGCTGAGAGAGAGATGCT 5580  
DB 5521 GCACAGGTGTCACCCCAATTTGTTGTGAGCAGCAGATGCTGAGAGAGAGATGCT 5580  
QY 5581 TCCATGCAATTTGSCAGATGTGTAGGACACCTGTGTGACCCGAGAGTGGGTGTGACA 5640  
DB 5581 TCCATGCAATTTGSCAGATGTGTAGGACACCTGTGTGACCCGAGAGTGGGTGTGACA 5640  
QY 5641 GTGATGACCTTACAGTCCAGAGAGTGTGACACCTGACCTGATACCCAGATCCCCACA 5700  
DB 5641 GTGATGACCTTACAGTCCAGAGAGTGTGACACCTGACCTGATACCCAGATCCCCACA 5700  
QY 5701 GCCACTACTGA 5711  
DB 5701 GCCACTACTGA 5711



Fri Jun 27 10:47:21 2003

Search completed: June 27, 2003, 01:01:50  
Job time : 14243.3 secs

us-09-734-672-3.rge



Fri Jun 27 10:47:47 2003

us-09-734-672-3\_copy\_2100\_2300.oli.rst

Cto T a + 2201

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 05:04:22 ; Search time 1769 Seconds  
(without alignments)  
1840.188 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_2100\_2300  
Perfect score: 201  
Sequence: 1 AGGCACAGCAGAAACTACA.....TTGTCATCTCAGCCTTCA 201

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_plo:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	181	90.0	602 10 BE043993	BE043993 hK83C01.x
2	150	74.6	1089 13 BM452288	BM452288 AGENCOURT
3	48	23.9	443 14 BQ308670	BQ308670 MRO-B7600
4	44	21.2	212 14 U25782	U25782 HSU25782 no
5	19	9.5	213 12 BG349328	BG349328 947028A03
6	19	9.5	219 9 A1414403	A1414403 ma399a01.x

7	19	9.5	387 12 BE367151	BE367151 RCI-GN002
8	19	9.5	412 10 AM531177	AM531177 UI-R-BH1-
9	19	9.5	444 12 BG349012	BG349012 947028A03
10	19	9.5	448 9 AA283425	AA283425 vb90e09.r
11	19	9.5	454 9 AA242194	AA242194 my30e01.r
12	19	9.5	493 14 BQ031604	BQ031604 UI-1-CFO-
13	19	9.5	514 9 AA178652	AA178652 ml18g03.r
14	19	9.5	533 13 BJ195365	BJ195365 BJ195365
15	19	9.5	563 13 BM045989	BM045989 603625142
16	19	9.5	580 9 A1504935	A1504935 v132c12.x
17	19	9.5	624 13 BJ205435	BJ205435 BJ205435
18	19	9.5	2287 11 BC020362	BC020362 Mus muscu
19	18	9.0	259 12 BE560517	BE560517 UI-R-CO-h
20	18	9.0	293 14 T50377	T50377 yb55c05.s1
21	18	9.0	309 14 T51068	T51068 yb55c05.r1
22	18	9.0	376 17 CNS0288Y	AT232315 Tetradon
23	18	9.0	399 17 A0166402	A0166402 HS_3128_B
24	18	9.0	443 17 B54094	B54094 CIT-HSP-201
25	18	9.0	476 10 BE096207	BE096207 UI-R-BU0-
26	18	9.0	486 17 BH513308	BH513308 BQGB12TR
27	18	9.0	512 14 B0130483	B0130483 i186h06.x
28	18	9.0	523 13 B1678186	B1678186 SMS866 SW
29	18	9.0	535 17 BH339312	BH339312 CH230-41D
30	18	9.0	540 17 A0947534	A0947534 Sheared D
31	18	9.0	544 9 AA964534	AA964534 UI-R-CO-g
32	18	9.0	554 17 TA291D05P	TA291D05P T. brucei
33	18	9.0	555 17 TA295E06P	AT484066 T. brucei
34	18	9.0	640 9 A0077451	A0077451 A0077451
35	18	9.0	642 13 B1087604	B1087604 602852095
36	18	9.0	660 17 A2220818	A2220818 Sheared D
37	18	9.0	715 17 BH176278	BH176278 005.N.10-
38	18	9.0	715 17 CNS07137	AT613237 T7 end of
39	18	9.0	723 17 AG181688	AG181688 Pan trogl
40	18	9.0	732 17 A2216722	A2216722 Sheared D
41	18	9.0	761 17 A0450104	A0450104 500008A05
42	18	9.0	837 17 A2199793	A2199793 SP_1040_B
43	18	9.0	862 17 A2677885	A2677885 EMT1H61TF
44	18	9.0	871 17 CANS059JH	AT327302 Tetradon
45	18	9.0	886 17 CANS0216T	AT196542 Tetradon

ALIGNMENTS

RESULT 1  
BE043993/c 602 bp mRNA linear EST 08-JUN-2000  
LOCUS hK83C01.x1 NCI CGAP Lym12 Homo sapiens cDNA clone IMAGE:3003264.3'  
DEFINITION similar to gb:U14680 BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN (HUMAN); mRNA sequence.  
ACCESSION BE043993  
VERSION BE043993.1 GI:8361046  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 602)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Life Technologies catalog #: 11547-015  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.lnl.gov  
Seq primer: -400p from Glibco  
High quality sequence stop: 418.  
Location/Qualifiers 1..602  
FEATURES  
SOURCE.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3003264"
/clone_lib="NCI CGAP Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/Note="Organ: lymph node; Vector: pCMV-SPORT6; Site:1:
Salt: Site:2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"

BASE COUNT      131 a      114 c      113 g      243 t      1 others
ORIGIN

Query Match      90.0%; Score 181; DB 10; Length 602;
Best Local Similarity 100.0%; Pred. No. 3.4e-84;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ACTCATGGAAGTAAAGAACTGCAAGTGGAGCCAGAGAGTAAAGCAATGACAA 80
    |||||||
DB 188 ACTCATGGAAGTAAAGAACTGCAAGTGGAGCCAGAGAGTAAAGCAATGACAA 129
    |||||||
QY 81 GACAGTAAAGACAGACAGTATCTTCCAGAGCTCAATTAACAATGACCTGG 140
    |||||||
DB 128 GACAGTAAAGACAGACAGTATCTTCCAGAGCTCAATTAACAATGACCTGG 69
    |||||||
QY 141 TTCCTTAACTAGTGTTCACAAATACAGTAACCTTAAGAATTGTCAATCCCTGCTCC 200
    |||||||
DB 68 TTCCTTAACTAGTGTTCACAAATACAGTAACCTTAAGAATTGTCAATCCCTGCTCC 9
    |||||||
QY 201 A 201
    |
DB 8 A 8

RESULT 2
BM452288      1089 bp      mRNA      linear      EST 05-FEB-2002
LOCUS          AGENCOURT_6386302 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526666
DEFINITION     5', mRNA sequence.
ACCESSION      BM452288
VERSION        BM452288.1 GI:16501328
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1089)
NIH-MGC http://mgi.mgi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rt@mail.nih.gov
Tissue Procurement: ATCC/DCRD/PTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12200 row: k column: 19
High quality sequence stop: 667.
Location/Qualifiers
1. 1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5526666"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life

```

```

Technologies."
BASE COUNT      389 a      212 c      248 g      240 t
ORIGIN

Query Match      74.6%; Score 150; DB 13; Length 1089;
Best Local Similarity 99.5%; Pred. No. 7.5e-68;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTTACACATCATGAGGTAAAGCACTGCAACTGGAGCCAAAG 60
    |||||||
DB 188 AGGCACAGCAGAACTTACACATCATGAGGTAAAGCACTGCAACTGGAGCCAAAG 247
    |||||||
QY 61 AGTAACAGCCAAATGACAGACAGTAAGTAAAGCATGATGATCTTCCAGAGCTG 120
    |||||||
DB 248 AGTAACAGCCAAATGACAGACAGTAAGTAAAGCATGATGATCTTCCAGAGCTG 307
    |||||||
QY 121 AAGTTAAACAATGACCTGCTTCTTACTTAAGTGTCAATTAACAGTAACCTTAAGA 180
    |||||||
DB 308 AAGTTAAACAATGACCTGCTTCTTACTTAAGTGTCAATTAACAGTAACCTTAAGA 367
    |||||||
QY 181 TTGTCAATCCCTGACCTTCCA 201
    |||||||
DB 368 TTGTCAATCCCTGACCTTCCA 388

RESULT 3
B0308670      443 bp      mRNA      linear      EST 16-MAY-2002
LOCUS          MRO-BR6000-020701-101-b03 BR6000 Homo sapiens cDNA, mRNA sequence.
DEFINITION     B0308670.1 GI:20850612
ACCESSION      B0308670
VERSION        B0308670.1
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 443)
Dias Neto,E., Garcia Correia,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR06t2=MR0-BR6000-
020701-101-b03&t3=2001-07-02&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 355.
Location/Qualifiers
1. 443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BR6000"
/dev_stage="Adult"
/Note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of low
mRNA and cDNA amplification were performed under low

```



High quality sequence stop: 206.

# FEATURES

Location/Qualifiers

1. 219  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:318792"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCCGCAATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a cot = 5. Library constructed by Bento Soares and M. Fatima Boudado. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## BASE COUNT

61 a 46 c 32 g 80 t

## ORIGIN

Query Match 9.5%; Score 19; DB 9; Length 219;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

143 CTTTACTAGTGTCAAA 161  
21 CTTTACTAGTGTCAAA 3

## Db

RESULT 7  
BF367151 387 bp mRNA linear EST 24-NOV-2000  
LOCUS RCI-GN0021-300800-022-cl1 GN0021 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF367151  
VERSION BF367151.1 GI:11329176  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 387)  
Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

Shogun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20702663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gehtml2.pl?cl=RC1&c2=RC1-GN0021-  
300800-022-cl1&c3=2000-08-30&c4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 387.

## FEATURES

Location/Qualifiers

1. 387  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0021"

/dev\_stage="Adult"

/note="Organ: placenta-normal; Vector: puc18; Site\_1: SmaI  
/ Site\_2: SmaI; A mini-library was made by cloning  
products derived from ONESIES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

## BASE COUNT

90 a 108 c 52 g 137 t

## ORIGIN

Query Match 9.5%; Score 19; DB 12; Length 387;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

128 CAATGCACTGCTCTTT 146  
186 CAATGCACTGCTCTTT 204

## Db

RESULT 8  
AM531177/c 412 bp mRNA linear EST 06-MAR-2000  
LOCUS UI-R-BT1-akv-e-12-0-UI.s1 UI-R-BT1 Rattus norvegicus cDNA clone  
DEFINITION UI-R-BT1-akv-e-12-0-UI 3', mRNA sequence.  
ACCESSION AM531177  
VERSION AM531177.1 GI:7173591  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 412)  
Boudado, M. F., Lennon, G. and Soares, M. B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: mscoares@iuiowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the Noli site  
and the oligo-dT track served to identify it as a clone from the  
normalized thalamus library cDNA library Preparation: M. B. Soares  
Lab Clone distribution: clones will be available through Research  
Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

source

Location/Qualifiers

1. 412  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BT1-akv-e-12-0-UI"  
/clone\_lib="UI-R-BT1"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The library  
UI-R-BT1 is a subtracted library derived from a mixture of  
the following tissues: hippocampus, thalamus, mid-brain,  
medulla, corpus striatum, cerebral cortex and testis. For  
a detailed description of the library from which this  
clone was derived, please visit our web site at  
ratest.eng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares; Genome Research  
6:791-806, 1996)  
TAG\_LIB=UT-R-BT1  
TAG\_TISSUE=thalamus  
TAG\_SEQ=CATCG\*

BASE COUNT 109 a 68 c 72 g 163 t  
ORIGIN

Query Match 9.5%; Score 19; DB 10; Length 412;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 TCTTTACTAGTGTCAA 160  
|||||  
Db 404 TCTTTACTAGTGTCAA 386

RESULT 9  
BG349012/c 444 bp mRNA linear EST 01-MAR-2001  
LOCUS 947028A03.x2 947 - 2 week shoot from Barkan lab Zea mays CDNA, mRNA  
DEFINITION  
SEQUENCE  
ACCESSION BG349012 GI:13177662  
VERSION BG349012.1  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
Walbot,V.  
1 (bases 1 to 444)

REFERENCE  
AUTHORS Walbot,V.  
TITLE Mize ESTs from various CDNA libraries sequenced at Stanford  
UNIVERSITY  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 723 8221  
Email: walbot@stanford.edu  
Plate: 947028 row: A column: 03.  
Location/Qualifiers

FEATURES  
source 1..444  
location/Qualifiers

/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="947 - 2 week shoot from Barkan lab"  
/issue\_type="leaf and stem, including leaf base"  
/dev\_stage="2 week old seedling (3 leaves)"  
/lab\_host="X11-Blue"  
/note="Organ: Shoot; Vector: Lambda ZAP (pBluescript SK-);  
Site\_1: EcoRI; Site\_2: XhoI; Directionally cloned using  
stratagene's unitap XR cDNA cloning kit with the 5' end  
at the EcoRI site. The library represents 8 x 10e5  
independent recombinant phage. The plants were greenhouse  
grown."

BASE COUNT 148 a 116 c 75 g 105 t  
ORIGIN

Query Match 9.5%; Score 19; DB 12; Length 444;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AACAGCCAAATGACAGA 82  
|||||  
Db 436 AACAGCCAAATGACAGA 418

RESULT 10  
AA285425 448 bp mRNA linear EST 08-APR-1997  
LOCUS

DEFINITION vb90e09.r1 Soares mouse 3nbms Mus musculus CDNA clone IMAGE:764296  
5', mRNA sequence.  
AA285425  
VERSION AA285425.1 GI:1929816  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Merra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:465216

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 432.  
Location/Qualifiers

1..448  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:764296"  
/clone\_lib="Soares mouse 3nbms"  
/sex="male"  
/issue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand CDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCGACGAGGAGCGCGCGCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded CDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p773 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through  
three rounds of normalization, and was constructed by  
Bento Soares and M.Fatima Bonaldo."

BASE COUNT 165 a 88 c 75 g 120 t  
ORIGIN

Query Match 9.5%; Score 19; DB 9; Length 448;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 TCTTTACTAGTGTCAA 160  
|||||  
Db 88 TCTTTACTAGTGTCAA 106

RESULT 11  
AA242194 454 bp mRNA linear EST 07-MAR-1997  
LOCUS my30e01.r1 Barstead mouse pooled organs MRLR84 Mus musculus CDNA  
DEFINITION clone IMAGE:697368 5', mRNA sequence.  
AA242194  
ACCESSION AA242194 GI:1873385  
VERSION AA242194.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:382284

Putative full length read  
vector to vector length is 635  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 480.  
Location/Qualifiers

FEATURES  
SOURCE

1. 514  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:621460"  
/clone\_lib="Soares mouse 3BDMs"  
/sex="male"  
/tissue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dt) primer [5',  
TGTACCAATCTGAGTGGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through  
three rounds of normalization, and was constructed by  
Bento Soares and M.Fatima Bonaldo."

BASE COUNT  
ORIGIN

Query Match 9.5%; Score 19; DB 9; Length 514;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 TCTTTTACTAGTGTCA 160  
|||||  
Db 161 TCTTTTACTAGTGTCA 179

RESULT 14 533 bp mRNA linear EST 24-JAN-2002  
LOCUS BJ195365/c BJ195365 normalized full length cDNA library, chloronemata,  
caulonemata and rhizoid-like protonemata Physcomitrella patens  
subsp. patens cDNA clone pphn23d17 5', mRNA sequence.  
ACCESSION BJ195365.1 GI:18363293  
VERSION BJ195365  
KEYWORDS EST.  
SOURCE Physcomitrella patens subsp. patens.  
ORGANISM Physcomitrella patens subsp. patens  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Bryophyta:  
Bryopsida: Funariidae: Funariales: Funariaceae: Physcomitrella:  
1 (bases 1 to 533)  
Fujiita,T., Suda,I.T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,  
'Carinci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe  
,M.

Comparison of the moss Physcomitrella patens genome with flowering  
plants genome  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo  
excised from a modified lps phage vector (Mo bi Tec, Germany). XhoI  
digested-5' end of cDNA is ligated to SalI site of the vector, and  
the BamHI digested-3' end including poly-A tail is ligated to BamHI  
site of the vector. cDNA insert could be amplified with  
conventional T7 and T3 primers. This normalized full-length cDNA  
library was generated basically according to the method described

in Genome Research 10, 1617-1630 (2000), Carinci, P. et al.  
Protonemata were blended by the POLYTRON, and then cultivated on  
the BCD medium containing 1um NAA (naphthalene acetic acid) for 8  
to 11 days under the continuous light.  
Location/Qualifiers

FEATURES  
SOURCE

1. 533  
/organism="Physcomitrella patens subsp. patens"  
/db\_xref="taxon:145481"  
/clone="pphn23d17"  
/clone\_lib="normalized full length cDNA library,  
chloronemata, caulonemata and rhizoid-like protonemata"  
/tissue\_type="mixture of chloronemata, caulonemata and  
rhizoid-like protonemata"  
BASE COUNT 121 a 110 c 162 g 140 t  
ORIGIN

Query Match 9.5%; Score 19; DB 13; Length 533;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CCAGAGACGATACACGCC 71  
|||||  
Db 239 CCAGAGAGATACACGCC 221

RESULT 15 563 bp mRNA linear EST 07-NOV-2001  
LOCUS BM045989 603625142P1 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:5451611 5',  
DEFINITION mRNA sequence.  
ACCESSION BM045989  
VERSION BM045989.1 GI:16775256  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
1 (bases 1 to 563)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs.fda.gov  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM1942 row: d column: 12  
High quality sequence stop: 389.  
Location/Qualifiers

1. 563  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5451611"  
/clone\_lib="NIH\_MGC\_40"  
/tissue\_type="Carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the Laboratory of Gerald M. Rubin (University  
of California, Berkeley) using Zap-CDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT 143 a 124 c 174 g 122 t  
ORIGIN

Query Match 9.5%; Score 19; DB 13; Length 563;  
Best Local Similarity 100.0%; Pred. No. 40;

Matches	19;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	75	TGACACAGACAGTAAAGA	93						
Db	410	TGACACAGACAGTAAAGA	428						

Search completed: June 27, 2003, 06:35:34  
 Job time : 1787 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 17:13:18 ; Search time 1905 Seconds

(without alignments)  
3070.690 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_2100\_2300

Perfect score: 201  
Sequence: 1 AGGCACAGCAGAACTACA.....TTGTCAATCTAGCCTTCA 201

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapept 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_cm:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hlg\_hum:\*

31: em\_hlg\_inv:\*

32: em\_hlg\_other:\*

33: em\_hlg\_mus:\*

34: em\_hlg\_pln:\*

35: em\_hlg\_rnd:\*

36: em\_hlg\_mam:\*

37: em\_hlg\_vit:\*

38: em\_sy:\*

39: em\_hlgo\_hum:\*

40: em\_hlgo\_mus:\*

41: em\_hlgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	5711	6 AR007333	AR007333 Sequence
2	201	100.0	5711	6 AR007335	AR007335 Sequence
3	201	100.0	5711	6 AR112808	AR112808 Sequence
4	201	100.0	5711	6 AR112809	AR112809 Sequence
5	201	100.0	5711	6 I59546	I59546 Sequence 1
6	150	74.6	298	9 AF110102	AF110102 Homo sapi
7	150	74.6	649	9 AF274503	AF274503 Homo sapi
8	150	74.6	4249	6 AR004691	AR004691 Sequence
9	150	74.6	4249	6 AR008177	AR008177 Sequence
10	150	74.6	4249	6 AR136960	AR136960 Sequence
11	150	74.6	4249	6 I76961	I76961 Sequence 21
12	150	74.6	4249	6 I80956	I80956 Sequence 21
13	150	74.6	4249	6 I81052	I81052 Sequence 21
14	150	74.6	5656	6 AR048657	AR048657 Sequence
15	150	74.6	5656	6 I40792	I40792 Sequence 1
16	150	74.6	5689	6 AR048659	AR048659 Sequence
17	150	74.6	5689	6 I40794	I40794 Sequence 3
18	150	74.6	5693	9 AF005068	AF005068 Homo sapi
19	150	74.6	5707	6 AR048667	AR048667 Sequence
20	150	74.6	5707	6 I40802	I40802 Sequence 11
21	150	74.6	5709	6 AR048658	AR048658 Sequence
22	150	74.6	5709	6 AR048663	AR048663 Sequence
23	150	74.6	5709	6 AR048664	AR048664 Sequence
24	150	74.6	5709	6 AR048665	AR048665 Sequence
25	150	74.6	5709	6 AX281859	AX281859 Sequence
26	150	74.6	5709	6 I40793	I40793 Sequence 2
27	150	74.6	5709	6 I40798	I40798 Sequence 7
28	150	74.6	5709	6 I40799	I40799 Sequence 8
29	150	74.6	5709	6 I40800	I40800 Sequence 9
30	150	74.6	5710	6 AR048662	AR048662 Sequence
31	150	74.6	5710	6 I40797	I40797 Sequence 6
32	150	74.6	5711	6 AR007334	AR007334 Sequence
33	150	74.6	5711	6 AR033056	AR033056 Sequence
34	150	74.6	5711	6 AR048660	AR048660 Sequence
35	150	74.6	5711	6 AR048666	AR048666 Sequence
36	150	74.6	5711	6 AR112810	AR112810 Sequence
37	150	74.6	5711	6 I40795	I40795 Sequence 4
38	150	74.6	5711	6 I40801	I40801 Sequence 10
39	150	74.6	5711	9 HSU14680	U14680 Homo sapien
40	150	74.6	5712	6 AR048668	AR048668 Sequence
41	150	74.6	5712	6 AR070223	AR070223 Sequence
42	150	74.6	5712	6 AR118507	AR118507 Sequence
43	150	74.6	5712	6 AR125601	AR125601 Sequence
44	150	74.6	5712	6 AR184044	AR184044 Sequence
45	150	74.6	5712	6 I40803	I40803 Sequence 12

#### ALIGNMENTS

RESULT 1  
LOCUS AR007333 5711 bp DNA linear PAT 04-DEC-1998  
DEFINITION Sequence 1 from patent US 5750400.  
ACCESSION AR007333  
VERSION AR007333.1 GI:3966817  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5711)  
AUTHORS Murphy,P.D., Allen,A.C., Alvaras,C.P., Crlitz,B.S., Olson,S.J.,  
Scheller,D.B., and Zeng,B.  
TITLE Coding sequences of the human BRCA1 gene  
JOURNAL Patent: US 5750400-A 1 12-MAY-1998;

FEATURES	Location/Qualifiers
SOURCE	1. 5711 /organism="unknown"
BASE COUNT	1953 a 1099 c 1277 g 1382 t
ORIGIN	
Query Match	100.0%: Score 201; DB 6; Length 5711;
Best Local Similarity	100.0%: Pred. No. 5.3e-103;
Matches 201; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGGCACAGCAGAAACCTACACATCATGGAAGGTAAAGAACTGCACACTGAGCCAGAG 60
DB	2100 AGGCACAGCAGAAACCTACACATCATGGAAGGTAAAGAACTGCACACTGAGCCAGAG 2159
OY	61 AGTAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGACTCTTCCAGAGCTG 120
DB	2160 AGTAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGACTCTTCCAGAGCTG 2219
OY	121 AAGTTAACAAATGCAACCTGGTTCTTTACTAGTGTCAATACCACTGAACTTAAAGAA 180
DB	2220 AAGTTAACAAATGCAACCTGGTTCTTTACTAGTGTCAATACCACTGAACTTAAAGAA 2279
OY	181 TTGTGCAATCCTAGCCTTCCA 201
DB	2280 TTGTGCAATCCTAGCCTTCCA 2300
RESULT 2	
AR007335	5711 bp DNA linear PAT 04-DEC-1998
LOCUS	AR007335
DEFINITION	Sequence 5 from patent US 5750400.
ACCESSION	AR007335
VERSION	AR007335.1 GI:3966819
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 5711)
AUTHORS	Murphy,P.D., Allen,A.C., Alvares,C.P., Critz,B.S., Olson,S.J., Schelter,D.B. and Zeng,B.
TITLE	Coding sequences of the human BRCA1 gene
JOURNAL	Patent: US 5750400-A 5 12-MAY-1998;
FEATURES	Location/Qualifiers
SOURCE	1. 5711 /organism="unknown"
BASE COUNT	1953 a 1098 c 1277 g 1383 t
ORIGIN	
Query Match	100.0%: Score 201; DB 6; Length 5711;
Best Local Similarity	100.0%: Pred. No. 5.3e-103;
Matches 201; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGGCACAGCAGAAACCTACACATCATGGAAGGTAAAGAACTGCACACTGAGCCAGAG 60
DB	2100 AGGCACAGCAGAAACCTACACATCATGGAAGGTAAAGAACTGCACACTGAGCCAGAG 2159
OY	61 AGTAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGACTCTTCCAGAGCTG 120
DB	2160 AGTAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGACTCTTCCAGAGCTG 2219
OY	121 AAGTTAACAAATGCAACCTGGTTCTTTACTAGTGTCAATACCACTGAACTTAAAGAA 180
DB	2220 AAGTTAACAAATGCAACCTGGTTCTTTACTAGTGTCAATACCACTGAACTTAAAGAA 2279
OY	181 TTGTGCAATCCTAGCCTTCCA 201
DB	2280 TTGTGCAATCCTAGCCTTCCA 2300
RESULT 3	
AR112808	5711 bp DNA linear PAT 16-MAY-2001
LOCUS	AR112808
DEFINITION	Sequence 1 from patent US 6130322.

ACCESSION	ARL12808
VERSION	ARL12808.1 GI:14092708
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 5711)
AUTHORS	Murphy,P.D., Allen,A.C.P., Alvares,C.P., Critz,B.S., Olson,S.J., Thuber,D. and Zeng,B.
JOURNAL	Coding sequences of the human BRCA1 gene
FEATURES	Patent: US 6130322-A 1 10-OCT-2000; Location/Qualifiers 1..5711
BASE COUNT	/organism="unknown"
ORIGIN	1953 a 1099 c 1277 g 1382 t
Query Match	100.0%; Score 201; DB 6; Length 5711;
Best Local Similarity	100.0%; Pred. No.5,3e-103;
Matches 201:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGGCACAGCAGAACCCTACAACCTCATGGAAGTAAAGAACCTCGCAACTGGAGCCAAAGANG 60
Db	2100 AGGCACAGCAGAACCCTACAACCTCATGGAAGTAAAGAACCTCGCAACTGGAGCCAAAGANG 2155
OY	61 AGTAACACGCCAATGATGACAGACAGCAAGTAAAAGACATACAGTAGTACTTTCCAGAGCTG 120
Db	2160 AGTAACACGCCAATGATGACAGACAGCAAGTAAAAGACATACAGTAGTACTTTCCAGAGCTG 2219
OY	121 AAGTTAACAAATGACACCTGGTGTTCTTTACTGAAGTGTCCAATKCCAGGTGAAGTAAAGAA 180
Db	2220 AAGTTAACAAATGACACCTGGTGTTCTTTACTGAAGTGTCCAATKCCAGGTGAAGTAAAGAA 2279
OY	181 TTTGTCAATCCTAGCCTTCOCA 201
Db	2280 TTTGTCAATCCTAGCCTTCOCA 2300
RESULT 4	
LOCUS	ARL12809 5711 bp DNA linear PAT 16-MAY-2001
DEFINITION	Sequence 3 from patent US 6130322.
ACCESSION	ARL12809
VERSION	ARL12809.1 GI:14092709
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 5711) Murphy,P.D., Allen,A.C.P., Alvares,C.P., Critz,B.S., Olson,S.J., Thuber,D. and Zeng,B.
TITLE	Coding sequences of the human BRCA1 gene
JOURNAL	Patent: US 6130322-A 3 10-OCT-2000;
FEATURES	Location/Qualifiers 1..5711
BASE COUNT	/organism="unknown"
ORIGIN	1953 a 1098 c 1277 g 1383 t
Query Match	100.0%; Score 201; DB 6; Length 5711;
Best Local Similarity	100.0%; Pred. No.5,3e-103;
Matches 201:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGGCACAGCAGAACCCTACAACCTCATGGAAGTAAAGAACCTCGCAACTGGAGCCAAAGANG 60
Db	2100 AGGCACAGCAGAACCCTACAACCTCATGGAAGTAAAGAACCTCGCAACTGGAGCCAAAGANG 2155
OY	61 AGTAACACGCCAATGATGACAGACAGCAAGTAAAAGACATACAGTAGTACTTTCCAGAGCTG 120
Db	2160 AGTAACACGCCAATGATGACAGACAGCAAGTAAAAGACATACAGTAGTACTTTCCAGAGCTG 2219
OY	121 AAGTTAACAAATGACACCTGGTGTTCTTTACTGAAGTGTCCAATKCCAGGTGAAGTAAAGAA 180
Db	2220 AAGTTAACAAATGACACCTGGTGTTCTTTACTGAAGTGTCCAATKCCAGGTGAAGTAAAGAA 2279

Db 2220 AAGTTACAAATGCACCTGGTCTTTACTAAGTTCATTAACAGTGAACCTTAAGCA 2279  
Oy 181 TTGTCAATCCTAGCCTTCCA 201  
|||||  
Db 2280 TTGTCAATCCTAGCCTTCCA 2300

RESULT 5  
LOCUS 159546 5711 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 1 from patent US 5654155.  
ACCESSION 159546 GI:2478178  
VERSION 159546.1  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5711)  
AUTHORS Murphy, P.D., Allen, A.C., Alvares, C.P., Critz, B.S., Olson, S.J.,  
Schelter, D.B., and Zeng, B.  
TITLE Consensus sequence of the human BRCA1 gene  
JOURNAL Patent: US 5654155-A 1 05-AUG-1997;  
FEATURES Location/Qualifiers  
source 1..5711  
BASE COUNT 1953 a 1099 c 1277 g 1382 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 5.3e-103;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGACCTGACACTGGAGCCAAAGAG 60  
|||||  
Db 2100 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGACCTGACACTGGAGCCAAAGAG 2159  
Oy 61 AGTAACAAGCCCAATGAAACAGACAGCAAGTAAGACATGACAGTACTTCCAGAGCTG 120  
|||||  
Db 2160 AGTAACAAGCCCAATGAAACAGACAGCAAGTAAGACATGACAGTACTTCCAGAGCTG 2219  
Oy 121 AAGTTACAATGCACTGCTTTCTTTACTAAGTTCATTAACAGTGAACCTTAAGAA 180  
|||||  
Db 2220 AAGTTACAATGCACTGCTTTCTTTACTAAGTTCATTAACAGTGAACCTTAAGAA 2279  
Oy 181 TTGTCAATCCTAGCCTTCCA 201  
|||||  
Db 2280 TTGTCAATCCTAGCCTTCCA 2300

RESULT 6  
AF110102 298 bp DNA linear PRI 08-JUN-1999  
LOCUS Homo sapiens breast and ovarian cancer susceptibility gene (BRCA1),  
DEFINITION exon 11 sequence.  
ACCESSION AF110102  
VERSION AF110102.1 GI:5006431  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 298)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE Corvello, C.M., Duarte, A.P.M., Mourao-Neto, M., and Simpson, A.J.G.  
JOURNAL Germline mutation at exon 11 of the BRCA1 gene in a female with  
family history of breast cancer  
REFERENCE 2 (bases 1 to 298)  
AUTHORS Unpublished  
JOURNAL Direct Submission  
TITLE Submitted (30-NOV-1998) Cancer Genetics, Ludwig Institute for  
Cancer Research, R. Prof Antonio Prudente, 105 - 4 andar, Sao Paulo  
01509-010, Brazil  
FEATURES Location/Qualifiers

source 1..298  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17q21"  
/cell\_type="Lymphocytes"  
gene <1..>298  
/gene="BRCA1"  
exon <1..>298  
/gene="BRCA1"  
misc\_feature 1..298  
/gene="BRCA1"  
/note="corresponds to breast and ovarian cancer  
susceptibility protein that is prematurely truncated due  
to mutation"  
variation 55  
/gene="BRCA1"  
/replace="g"  
BASE COUNT 119 a 60 c 56 g 63 t  
ORIGIN

Query Match 74.6%; Score 150; DB 9; Length 298;  
Best Local Similarity 99.5%; Pred. No. 5.8e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGACCTGACACTGGAGCCAAAGAG 60  
|||||  
Db 91 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGACCTGACACTGGAGCCAAAGAG 150  
Oy 61 AGTAACAAGCCCAATGAAACAGACAGCAAGTAAGACATGACAGTACTTCCAGAGCTG 120  
|||||  
Db 151 AGTAACAAGCCCAATGAAACAGACAGCAAGTAAGACATGACAGTACTTCCAGAGCTG 210  
Oy 121 AAGTTACAATGCACTGCTTTCTTTACTAAGTTCATTAACAGTGAACCTTAAGAA 180  
|||||  
Db 211 AAGTTACAATGCACTGCTTTCTTTACTAAGTTCATTAACAGTGAACCTTAAGAA 270  
Oy 181 TTGTCAATCCTAGCCTTCCA 201  
|||||  
Db 271 TTGTCAATCCTAGCCTTCCA 291

RESULT 7  
AF274503 649 bp mRNA linear PRI 06-JUN-2002  
LOCUS Homo sapiens breast and ovarian cancer susceptibility (BRCA1)  
DEFINITION pseudogene, partial mRNA sequence.  
ACCESSION AF274503  
VERSION AF274503.1 GI:8745572  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 649)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE Yassae, V.R., Zeinali, S., Harlirchi, I., Jarvandi, S.,  
Monaeghni, M.A., Hornby, D.P., and Dalton, A.  
JOURNAL Novel mutations in the BRCA1 and BRCA2 genes in Iranian women with  
early-onset breast cancer  
REFERENCE 2 (bases 1 to 649)  
AUTHORS Breast Cancer Res. 4 (4), R6 (2002)  
MEDLINE 22094890  
PUBMED 12100744  
JOURNAL Direct Submission  
TITLE Submitted (02-JUN-2000) Molecular Biology and Biotechnology, The  
University of Sheffield, D27, Firth Court, Western Bank, Sheffield,  
South Yorkshire S10 2TN, UK  
FEATURES Location/Qualifiers  
source 1..649  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

gene

<1..>649  
/gene="BRCA1"  
/note="breast and ovarian cancer susceptibility; mutant  
form; similar to Homo sapiens BRCA1 in Genbank Accession  
Number U14680"  
/pseudo

BASE COUNT 253 a 112 c 141 g 143 t  
ORIGIN

Query Match 74.6%: Score 150; DB 9; Length 649;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 60  
DB 85 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 144  
QY 61 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 120  
DB 145 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 204  
QY 121 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 180  
DB 205 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 264  
QY 181 TTTGTCAATCTAGCCTTCCA 201  
DB 265 TTTGTCAATCTAGCCTTCCA 285

RESULT 8  
AR004691 AR004691 4249 bp DNA linear PAT 04-DEC-1998  
LOCUS Sequence 21 from patent US 5747282.  
DEFINITION AR004691  
ACCESSION AR004691.1 GI:3965570  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 4249)  
AUTHORS Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
Harshman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.  
17Q-linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5747282-A 21 05-MAY-1998;  
FEATURES Location/Qualifiers  
source 1..4249  
/organism="unknown"

BASE COUNT 1480 a 739 c 908 g 1122 t  
ORIGIN

Query Match 74.6%: Score 150; DB 6; Length 4249;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 60  
DB 1512 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 1571  
QY 61 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 120  
DB 1572 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 1631  
QY 121 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 180  
DB 1632 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 1691  
QY 181 TTTGTCAATCTAGCCTTCCA 201  
DB 1692 TTTGTCAATCTAGCCTTCCA 1712

RESULT 9  
AR008177 AR008177 4249 bp DNA linear PAT 04-DEC-1998  
LOCUS Sequence 21 from patent US 5753441.  
DEFINITION AR008177  
ACCESSION AR008177.1 GI:3967286  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 4249)  
AUTHORS Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
Harshman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.  
17Q-linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5753441-A 21 19-MAY-1998;  
FEATURES Location/Qualifiers  
source 1..4249  
/organism="unknown"

BASE COUNT 1480 a 739 c 908 g 1122 t  
ORIGIN

Query Match 74.6%: Score 150; DB 6; Length 4249;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 60  
DB 1512 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 1571  
QY 61 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 120  
DB 1572 AGTAACAGCCCAATATACAGACAGCAAGTAAAGACATGACAGTACTTTCCAGAGCTG 1631  
QY 121 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 180  
DB 1632 AAGTTACAAATGACACCTGTTCTTTACTAGTGTTCAAATACAGTGAATTAAGAA 1691  
QY 181 TTTGTCAATCTAGCCTTCCA 201  
DB 1692 TTTGTCAATCTAGCCTTCCA 1712

RESULT 10  
AR136960 AR136960 4249 bp DNA linear PAT 16-JUN-2001  
LOCUS Sequence 21 from patent US 6162897.  
DEFINITION AR136960  
ACCESSION AR136960.1 GI:14478210  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 4249)  
AUTHORS Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
Harshman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.  
17Q-linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 6162897-A 21 19-DEC-2000;  
FEATURES Location/Qualifiers  
source 1..4249  
/organism="unknown"

BASE COUNT 1480 a 739 c 908 g 1122 t  
ORIGIN

Query Match 74.6%: Score 150; DB 6; Length 4249;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 60  
DB 1512 AGGCACAGCAGAAACCTACACTCATGAGAGTAAGAACCTGCACACTGGAGCCAGAAG 1571

Qy	61	AGTAAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGATCTTCCAGAGCTG	120
Db	1572	AGTAAACAAGCCAAATGAAACAGACAGTAAGTAAAGACATGACATGATCTTCCAGAGCTG	1631
Qy	121	AAGTTAACAAATGACACCTGTTCTTTACTAAGTTCAAATACCAAGTGAACCTAAAGAA	180
Db	1632	AAGTTAACAAATGACACCTGTTCTTTACTAAGTTCAAATACCAAGTGAACCTAAAGAA	1691
Qy	181	TTTGCAATCCTAGCCTTCCA	201
Db	1692	TTTGCAATCCTAGCCTTCCA	1712
RESULT	11		
LOCUS	176961	4249 bp	DNA
DEFINITION	Sequence 21 from patent US 5693473.	linear	PAT 03-APR-1998
ACCESSION	176961		
VERSION	176961.1	GI:3013115	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 4249)		
AUTHORS	Shattuck-Eidens,D.M., Simard,J., Durocher,F., Eml,M. and Nakamura,Y.		
TITLE	Linked breast and ovarian cancer susceptibility gene		
JOURNAL	Patent: US 5693473-A 21 02-DEC-1997;		
FEATURES	Location/Qualifiers		
SOURCE	1..4249		
BASE COUNT	1480 a	739 c	908 g 1122 t
ORIGIN	/organism="unknown"		

Query Match	74.6%: Score 150; DB 6; Length 4249;
Best Local Similarity	99.5%: Pred. No. 5.1e-74;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 1 AGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCACACTGAGAGCCAGAG 60	
Db 1512 AGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCACACTGAGAGCCAGAG 15711	
QY 61 AGTAAACAAGCCAAATGAAACAGACAAGTAAAGACATGACAGTACTTTTCCAGAGCTG 120	
Db 1572 AGTAAACAAGCCAAATGAAACAGACAAGTAAAGACATGACAGGATACTTTCCAGAGCTG 16312	
QY 121 AAGTTAAACAATGAGACCTGGTTCTTTACTAGTGTTCAAATACACAGAGCACTTAAAGAA 180	
Db 1632 AAGTTAAACAATGAGACCTGGTTCTTTACTAGTGTTCAAATACACAGAGCACTTAAAGAA 1691	
QY 181 TTGTGCAATCCTAGCCTTCCA 201	
Db 1692 TTGTGCAATCCTAGCCTTCCA 1712	
RESULT 12	
LOCUS 180956 4249 bp DNA Linear PAT 10-JUN-1998	
DEFINITION Sequence 21 from patent US 5709999.	
ACCESSION 180956	
VERSION 180956.1 GI:3209246	
KEYWORDS	
SOURCE Unknown.	
ORGANISM Unknown.	
REFERENCE 1 (bases 1 to 4249)	
AUTHORS Shattuck-Eidens,D.M., Simard,J., Durocher,F., Emi,M. and Nakamura,Y.	
TITLE Linked breast and ovarian cancer susceptibility gene	
JOURNAL Patent: US 5709999-A 21 20-JUN-1998;	
FEATURES location/qualifiers	
Source 1..4249	
/organism="unknown"	

	BASE COUNT	1480 a	739 c	908 g	1122 t	
ORIGIN						
Query Match			74.6%;	Score 150;	DB 6;	Length 4249;
Best Local Similarity			99.5%;	Pred. No. 5.1e-74;		
Matches	200;	Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
QY	1	AGGCACACAGAAAACTCTACACTCATGTGAAGAGTAAAGAACTCGCACTGGAGCCAAAGAG	60			
DB	1512	AGGCACACAGAAAACTCTCAACTCTGAGAGGTAAAGAACTGCAACTGGAGCCAAAGAG	1517			
QY	61	AGTAACAAGCCAAATGAACAGACAAGTAAAGACATGCACGTGACTTCTCCAGAGCTG	120			
DB	1572	AGTACAAGCCAAATGAACAGACAAGTAAAGACATGCACGTGACTTCTCCAGAGCTG	1631			
QY	121	AAGTTAACAAATGCACCTCGTTCCTTTACTAAGTGTCTAAATTACCAGTGAACCTTAAGAA	180			
DB	1632	AAGTTAACAAATGCACTCGTTCCTTTACTAAGTGTCTAAATTACCAGTGAACCTTAAGAA	1691			
QY	181	TTTGTCAATCTAGACCTTCCA	201			
DB	1692	TTTGTCAATCTAGACCTTCCA	1712			

LOCUS	181052	4249 bp	DNA	linear	PAT 10-JUN-1998
DEFINITION	Sequence 21 from patent US 5710001.				
ACCESSION	181052				
VERSION	181052.1	GI:3209342			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 4249) Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A., Hartsham,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W. and Futreal,P.Andrew. 17q-linked breast and ovarian cancer susceptibility gene Patent: US 5710001-A 21 20-JAN-1998; Location/Qualifiers 1..4249 source				
TITLE					
JOURNAL					
FEATURES					
BASE COUNT	1480 a	739 c	908 g	1122 t	
ORIGIN					
Query Match	74.6%: Score 150; DB 6; Length 4249;				
Best Local Similarity	99.5%: Pred. No.5.1e-74;				
Matches 200; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
QY	1 AGGCACAGCAGAAACCTCACAACATCATGTGAAGGTAAAGAACTGCACACTGGAGCCAGAG 60				
DB	1512 AGGCACAGCAGAAACCTCACAACATCATGTGAAGGTAAAGAACTGCACACTGGAGCCAGAG 1571				
QY	61 AGTAAACAAGCCAAATGAACAGACAGAGTAAAGACATGACAGTGAATCTTTCCAGACCTG 120				
DB	1572 AGTAAACAAGCCAAATGAACAGACAGAGTAAAGACATGACAGTGAATCTTTCCAGACCTG 1631				
QY	121 AAGTTAAACAATGACACCTGGTCTTTTACTAAGTGCTTCAATACCAGTGAAGCTTAAAGAA 180				
DB	1632 AAGTTAAACAATGACACCTGGTCTTTTACTAAGTGCTTCAATACCAGTGAAGCTTAAAGAA 1691				
QY	181 TTGTGCATCTCAGCTTCCA 201				
DB	1692 TTGTGCATCTCAGCTTCCA 1712				
RESULT 14					
LOCUS	AR048657	5656 bp	DNA	linear	PAT 29-SEP-1999
DEFINITION	Sequence 1 from patent US 5821328.				
ACCESSION	AR048657				
VERSION	AR048657.1	GI:5971000			

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 5656)  
TITLE King, M.-C., Friedman, L., Ostermeyer, B., Rowell, S., Lynch, E., Szabo, C., and Lee, M.  
JOURNAL Genetic markers for breast, ovarian, and prostatic cancer  
PATENT: US 5821328-A 1 13-OCT-1998;  
FEATURES Location/Qualifiers  
source 1..5656  
/organism="unknown"  
BASE COUNT 1939 a 1087 c 1263 g 1367 t  
ORIGIN

Query Match 74.6%: Score 150; DB 6; Length 5656;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTTACACTCATGAGTAAGAGTAAGAACTGCAACTGGAGCCAGAAG 60  
DB 2046 AGGCACAGCAGAACTTACACTCATGAGTAAGAGTAAGAACTGCAACTGGAGCCAGAAG 2105  
QY 61 AGTAACAAGCCAAATGACAGACAGTAAGACATGACATGATATCTTTCCAGAGCTG 120  
DB 2106 AGTAACAAGCCAAATGACAGACAGTAAGACATGACATGATATCTTTCCAGAGCTG 2165  
QY 121 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATACCACTGACTTAAGAA 180  
DB 2166 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATACCACTGACTTAAGAA 2225  
QY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2226 TTTGTCAATCCTAGCCTTCCA 2246

RESULT 15  
LOCUS 140792 5656 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 1 from patent US 5622829.  
ACCESSION 140792  
VERSION 140792.1 GI:2082272  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 5656)  
TITLE King, M.-C., Friedman, L., Ostermeyer, B., Rowell, S., Lynch, E., Szabo, C., and Lee, M.  
JOURNAL Genetic markers for breast, ovarian, and prostatic cancer  
PATENT: US 5622829-A 1 22-APR-1997;  
FEATURES Location/Qualifiers  
source 1..5656  
/organism="unknown"  
BASE COUNT 1939 a 1087 c 1263 g 1367 t  
ORIGIN

Query Match 74.6%: Score 150; DB 6; Length 5656;  
Best Local Similarity 99.5%: Pred. No. 5.1e-74;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTTACACTCATGAGTAAGAGTAAGAACTGCAACTGGAGCCAGAAG 60  
DB 2046 AGGCACAGCAGAACTTACACTCATGAGTAAGAGTAAGAACTGCAACTGGAGCCAGAAG 2105  
QY 61 AGTAACAAGCCAAATGACAGACAGTAAGACATGACATGATATCTTTCCAGAGCTG 120  
DB 2106 AGTAACAAGCCAAATGACAGACAGTAAGACATGACATGATATCTTTCCAGAGCTG 2165  
QY 121 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATACCACTGACTTAAGAA 180  
DB 2166 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATACCACTGACTTAAGAA 2225

QY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2226 TTTGTCAATCCTAGCCTTCCA 2246

Search completed: June 27, 2003, 06:01:23  
Job time: 1929 secs





Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
1	AGGCACAGCAAACTCTCAACATCTGAGAGTAAAGAACTGCAACTGAGCCAGAAG	60				
2100	AGGCACAGCAAACTCTCAACATCTGAGAGTAAAGAACTGCAACTGAGCCAGAAG	2159				
61	AGTAAACAAGCAAAATGAACAGACAAGTAAAGACATGACAGTGAATCTTCCAGAGCTG	120				
2160	AGTAAACAAGCAAAATGAACAGACAAGTAAAGACATGACAGTGAATCTTCCAGAGCTG	2219				
121	AAAGTAAACAAGTGAACCTGCTTTCTTTTCTAAGTGTTCGAATATACAGTGAAGCTTAAAGAA	180				
2220	AAAGTAAACAAGTGAACCTGCTTTCTTTTCTAAGTGTTCGAATATACAGTGAAGCTTAAAGAA	2279				
181	TTTGTCAATCTCTAGCCTTCCA	201				
2280	TTTGTCAATCTCTAGCCTTCCA	2300				

ID	AAV62180	AAV62180 standard; DNA: 5711 BP.
XX	AAV62180;	
XX	11-FEB-1999 (first entry)	
XX	BRCA1 (om11) coding sequence.	
XX	BRCA1; mutation detection; disease screening; multiple allele variation;	
KW	breast cancer; ovarian cancer; cystic fibrosis; Li-Fraumeni syndrome;	
KW	Duchenne muscular dystrophy; Becker muscular dystrophy; ss.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	120..5711
XX		/*tag= a
XX	W09844157-A2.	
XX	08-OCT-1998.	
PD		
XX	26-MAR-1998; 98WO-US06002.	
PF		
XX	28-MAR-1997; 97US-0825487.	
PR		
XX	(ONCO-) ONCOMED INC.	
PA		
XX	Murphy PD, White MB;	
PI		
XX	WPI: 1998-542713/46.	
DR	P-PSDB; AAM79665.	
XX		
PT	Identifying variations in polynucleotide sequences - using allele	
PT	specific hybridisation assay, sequence variation locating assay, and	
XX	direct sequencing, in a stepwise procedure	
PS	Disclosure: Fig 1a-j; 62pp; English.	
XX		
XX	This sequence encodes the human BRCA (om11) protein, and was used to test	
CC	the method of the invention. The method is for determining the presence	
CC	or absence of a sequence variation in a gene sample, and comprises:	
CC	(a) performing an allele specific hybridisation assay for one or more	
CC	pre-determined sequence variations; (b) if no pre-determined sequence	
CC	variation found in step (a) then performing a sequence variation location	
CC	assay; (c1) if no sequence variation found in step (b) then sequencing	
CC	the gene sample; (c11) if sequence variation is found in step (b) then	
CC	targeted confirmatory sequencing is performed; and (d) determining the	
CC	presence of a sequence variation by analysing the sequence(s) obtained in	
CC	step (c1) or step (c11) against a reference sample. Alternatively, step	
CC	(a) or step (b) is omitted from the method. The invention provides a	
CC	stepwise and integrated method for the efficient and accurate detection	
CC	of variations in polynucleotide sequences, being directed towards	
CC	screening for diseases associated with multiple allele variations,	
CC	including breast and ovarian cancer, cystic fibrosis, Duchenne and Becker	
CC	muscular dystrophy, and Li-Fraumeni syndrome.	
XX		
XX	Sequence 5711 BP; 1953 A; 1099 C; 1277 G; 1382 T; 0 other;	
XX		
Query Match	100.0%; Score 201; DB 19; Length 5711;	
Best Local Similarity	100.0%; Pred. No. 6,4e-97;	
Matches 201; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 AGGCACAGCAGAAACCTACACTCATGTGAGAGCTAAGACCTGCACTGAGCAGCAAGAG	60
Db	2100 AGGCACAGCAGAAACCTACACTCATGTGAGAGCTAAGACCTGCACTGAGCAGCAAGAG	2155
OY	61 AGTACAGACCCAAATGAGACAGACAGTAAGAGACATGACAGTATCTTTCCAGAGCTG	120

ID	AAV62180	AAV62180 standard; DNA: 5711 BP.
XX	AAV62180;	
AC	11-FEB-1999 (first entry)	
DT	BRCA1 (om11) coding sequence.	
XX		
DE	BRCA1; mutation detection; disease screening; multiple allele variation;	
XX	breast cancer; ovarian cancer; cystic fibrosis; Li-Fraumeni syndrome;	
KW	Duchenne muscular dystrophy; Becker muscular dystrophy; ss.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	120..5711
XX		/*tag= a
XX	W09844157-A2.	
PD	08-OCT-1998.	
PF	26-MAR-1998; 98WO-US06002.	
XX		
PR	28-MAR-1997; 97US-0825487.	
XX		
PA	(ONCO-) ONCOMED INC.	
XX		
PI	Murphy PD, White MB;	
XX		
DR	WPI: 1998-542713/46.	
XX	P-PSDB; AAM79665.	
PT	Identifying variations in polynucleotide sequences - using allele	
PT	specific hybridisation assay, sequence variation locating assay, and	
XX	direct sequencing, in a stepwise procedure	
PS	Disclosure: Fig 1a-j; 62pp; English.	
XX		
CC	This sequence encodes the human BRCA (om11) protein, and was used to test	
CC	the method of the invention. The method is for determining the presence	
CC	or absence of a sequence variation in a gene sample, and comprises:	
CC	(a) performing an allele specific hybridisation assay for one or more	
CC	pre-determined sequence variations; (b) if no pre-determined sequence	
CC	variation found in step (a) then performing a sequence variation location	
CC	assay; (c1) if no sequence variation found in step (b) then sequencing	
CC	the gene sample; (c11) if sequence variation is found in step (b) then	
CC	targeted confirmatory sequencing is performed; and (d) determining the	
CC	presence of a sequence variation by analysing the sequence(s) obtained in	
CC	step (c1) or step (c11) against a reference sample. Alternatively, step	
CC	(a) or step (b) is omitted from the method. The invention provides a	
CC	stepwise and integrated method for the efficient and accurate detection	
CC	of variations in polynucleotide sequences, being directed towards	
CC	screening for diseases associated with multiple allele variations,	
CC	including breast and ovarian cancer, cystic fibrosis, Duchenne and Becker	
CC	muscular dystrophy, and Li-Fraumeni syndrome.	
XX		
XX	Sequence 5711 BP; 1953 A; 1099 C; 1277 G; 1382 T; 0 other;	
XX		
QY	Query Match	100.0%; Score 201; DB 19; Length 5711;
XX	Best Local Similarity	100.0%; Pred. No. 6,4e-97;
Db	Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
1	AGGCACAGCAGAAACCTACACTCATGTGAGAGCTAAGACCTGCACTGAGCAGCAGANG	60
2100	AGGCACAGCAGAAACCTACACTCATGTGAGAGCTAAGACCTGCACTGAGCAGCAGANG	2155
61	AGTACAGACCCAAATGAGACAGACAGTAAAGACATGACAGTATCTTTCCAGAGCTG	120

Db 2160 AGTAACAAGCCAAATGAACAGACAGTAAGACATGACATGATCTTCCAGAGCTG 2219  
 QY 121 AGTTACCAATGACCTGGTCTTTACTAGTGTCAATACCACTGAACTTAAGAA 180  
 Db 2220 AAGTTACCAATGACCTGGTCTTTACTAGTGTCAATACCACTGAACTTAAGAA 2279  
 QY 181 TTGTCAATCTCTAGCCTTCCA 201  
 Db 2280 TTGTCAATCTCTAGCCTTCCA 2300

RESULT 3  
 AAV46466 ID AAV46466 standard; cDNA; 5711 BP.  
 AC AAV46466;  
 XX  
 DT 18-NOV-1998 (first entry)  
 XX  
 DE Human BRCA1 omi3 polymorphism #2 cDNA.  
 XX  
 KW BRCA1; omi3; human; breast and ovarian cancer predisposing gene;  
 KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 KM chromosome 17q; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= "BRCA1 omi3 protein"  
 FT 2430  
 FT variation  
 FT /\*tag= b  
 FT /note= "This polymorphic variation can be a T or C  
 nucleotide"  
 FT  
 FT  
 XX  
 XX US5750400-A.  
 XX  
 XX 12-MAY-1998.  
 XX  
 XX 12-FEB-1997; 97US-0798691.  
 XX  
 XX 12-FEB-1996; 96US-0598591.  
 PR 12-FEB-1997; 97US-0798691.  
 XX  
 XX (ONCO-) ONCOMED INC.  
 PA  
 XX Allen AC, Alvaras CP, Critz BS, Murphy PD, Olson SJ;  
 PI Schelter DB, Zeng B;  
 PI  
 XX  
 XX WPI; 1998-296774/26.  
 XX  
 XX BRCA1 omi gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 PT  
 XX  
 PS Claim 2e; Page -: 54pp; English.  
 XX  
 XX This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) omi3 gene in which a polymorphic variation occurs at  
 CC nucleotide 2430. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 omi3 gene represented in AAV46450.  
 CC  
 XX

SO Sequence 5711 BP; 1953 A; 1097 C; 1277 G; 1383 T; 1 other;  
 Query Match 100.0%; Score 201; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 6,4e-97;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCACAGCAAGAACTACACATCATGAGGTAAGAACTGCACTGAGCCAGAAG 60  
 Db 2100 AGGCACAGCAAGAACTACACATCATGAGGTAAGAACTGCACTGAGCCAGAAG 2159  
 QY 61 AGTAACAAGCCAAATGAACAGACAGTAAGACATGACATGATCTTCCAGAGCTG 120  
 Db 2160 AGTAACAAGCCAAATGAACAGACAGTAAGACATGACATGATCTTCCAGAGCTG 2219  
 QY 121 AAGTTACCAATGACCTGGTCTTTACTAGTGTCAATACCACTGAACTTAAGAA 180  
 Db 2220 AAGTTACCAATGACCTGGTCTTTACTAGTGTCAATACCACTGAACTTAAGAA 2279  
 QY 181 TTGTCAATCTCTAGCCTTCCA 201  
 Db 2280 TTGTCAATCTCTAGCCTTCCA 2300

RESULT 4  
 AAV46467 ID AAV46467 standard; cDNA; 5711 BP.  
 AC AAV46467;  
 XX  
 DT 18-NOV-1998 (first entry)  
 XX  
 DE Human BRCA1 omi3 polymorphism #3 cDNA.  
 XX  
 KW BRCA1; omi3; human; breast and ovarian cancer predisposing gene;  
 KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 KM chromosome 17q; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= "BRCA1 omi3 protein"  
 FT 2731  
 FT variation  
 FT /\*tag= b  
 FT /note= "This polymorphic variation can be a C or T  
 nucleotide"  
 FT  
 FT  
 XX  
 XX US5750400-A.  
 XX  
 XX 12-MAY-1998.  
 XX  
 XX 12-FEB-1997; 97US-0798691.  
 XX  
 XX 12-FEB-1996; 96US-0598591.  
 PR 12-FEB-1997; 97US-0798691.  
 XX  
 XX (ONCO-) ONCOMED INC.  
 PA  
 XX Allen AC, Alvaras CP, Critz BS, Murphy PD, Olson SJ;  
 PI Schelter DB, Zeng B;  
 PI  
 XX  
 XX WPI; 1998-296774/26.  
 XX  
 XX BRCA1 omi gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 PT  
 XX  
 PS Claim 2e; Page -: 54pp; English.  
 XX  
 XX This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) omi3 gene in which a polymorphic variation occurs at  
 CC nucleotide 2731. This sequence and other polymorphic variations of this  
 CC

CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers,  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 omi3 gene represented in AAV46450.

CC Sequence 5711 BP; 1953 A; 1098 C; 1277 G; 1382 T; 1 other;

Query Match 100.0%; Score 201; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCACAGCAGAAACCTACACTCATGAGTAAGAACTGCAACTGGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAAACCTACACTCATGAGTAAGAACTGCAACTGGAGCCAGAG 2159  
OY 61 AGTACACAGCCAAATGAGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTACACAGCCAAATGAGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 2219  
OY 121 AAGTTAACAAATGACACCTGCTCTTTACTAGTGTCAATACAGTGAAGTAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTCTTTACTAGTGTCAATACAGTGAAGTAAGAA 2279  
OY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2280 TTTGTCAATCCTAGCCTTCCA 2300

RESULT 5  
AAV46468  
ID AAV46468 standard; cDNA; 5711 BP.

AC AAV46468;  
XX  
DT 18-NOV-1998 (first entry)

XX Human BRCA1 omi3 polymorphism #4 cDNA.

XX BRCA1: omi3; human: breast and ovarian cancer predisposing gene;  
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH 120..5711  
FT CDS

FT variation  
FT /tag= a "BRCA1 omi3 protein"  
FT 3232  
FT /note= "This polymorphic variation can be an A or G  
nucleotide"

XX US5750400-A.

XX 12-MAY-1998.

XX 12-FEB-1997; 97US-0798691.

XX 12-FEB-1996; 96US-0598591.

XX 12-FEB-1997; 97US-0798691.

XX (ONCO-) ONCORMED INC.

XX Allen AC, Alvares CP, Ciltz BS, Murphy PD, Olson SJ;  
PI Scheller DB, Zeng B;

XX WPI: 1998-296774/26.

XX BRCA1 omi gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer

PS Claim 2e; Page -: 54pp; English.

CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
CC predisposing gene) omi3 gene in which a polymorphic variation occurs at  
CC nucleotide 3232. This sequence and other polymorphic variations of this  
CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers,  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 omi3 gene represented in AAV46450.

CC Sequence 5711 BP; 1953 A; 1098 C; 1276 G; 1383 T; 1 other;

Query Match 100.0%; Score 201; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCACAGCAGAAACCTACACTCATGAGTAAGAACTGCAACTGGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAAACCTACACTCATGAGTAAGAACTGCAACTGGAGCCAGAG 2159  
OY 61 AGTACACAGCCAAATGAGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTACACAGCCAAATGAGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 2219  
OY 121 AAGTTAACAAATGACACCTGCTCTTTACTAGTGTCAATACAGTGAAGTAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTCTTTACTAGTGTCAATACAGTGAAGTAAGAA 2279  
OY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2280 TTTGTCAATCCTAGCCTTCCA 2300

RESULT 6  
AAV46469  
ID AAV46469 standard; cDNA; 5711 BP.

XX AAV46469;

XX 18-NOV-1998 (first entry)

XX Human BRCA1 omi3 polymorphism #5 cDNA.

XX BRCA1: omi2; human: breast and ovarian cancer predisposing gene;  
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 120..5711

XX variation

XX /tag= a "BRCA1 omi3 protein"  
XX 3667  
XX /note= "This polymorphic variation can be an A or G  
nucleotide"

XX US5750400-A.

XX 12-MAY-1998.  
 PD 12-FEB-1997; 97US-0798691.  
 XX 12-FEB-1997; 97US-0798691.  
 PE 12-FEB-1996; 96US-0598591.  
 PR 12-FEB-1997; 97US-0798691.  
 XX (ONCO-) ONCORMED INC.  
 PA Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
 PI Scheller DB, Zeng B;  
 PI WPI: 1998-296774/26.  
 DR BRCAl om1 gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 PS Claim 2e; Page -: 54pp; English.  
 XX This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) om13 gene in which a polymorphic variation occurs at  
 CC nucleotide 3667. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 om13 gene represented in AAV46450.  
 CC  
 CC Sequence 5711 BP; 1953 A; 1098 C; 1276 G; 1383 T; 1 other;  
 SQ  
 Query Match 100.0%; Score 201; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGCACAGCAGAAACCTACACTCATGGAAGGTAAAGACCTGCACTGAGCCAAAG 60  
 Db 2100 AGGCACAGCAGAAACCTACACTCATGGAAGGTAAAGACCTGCACTGAGCCAAAG 2159  
 QY 61 AGTACACAGCCAAATGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 120  
 Db 2160 AGTACACAGCCAAATGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 2219  
 QY 121 AAGTTAACAATGACACCTGCTCTTTACTAAGTGTCAATGCACTGAAGTTAAAGAA 180  
 Db 2220 AAGTTAACAATGACACCTGCTCTTTACTAAGTGTCAATGCACTGAAGTTAAAGAA 2279  
 QY 181 TTGTCAATCTAGCCTTCCA 201  
 Db 2280 TTGTCAATCTAGCCTTCCA 2300  
 RESULT 7  
 AAV46470  
 ID AAV46470 standard; cDNA; 5711 BP.  
 XX AAV46470;  
 AC  
 XX 18-NOV-1998 (first entry)  
 DT  
 XX Human BRCA1 om13 polymorphism #6 cDNA.  
 DE  
 XX BRCA1: om13: human: breast and ovarian cancer predisposing gene;  
 KW polymorphism: susceptibility; anti-oncogene; tumour suppressor;  
 KW chromosome 17q; ss.  
 XX

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= "BRCA1 om13 protein"  
 FT 4427  
 FT variation /\*tag= b  
 FT /note= "this polymorphic variation can be a T or C  
 FT nucleotide"  
 FT  
 FT US5750400-A.  
 PN 12-MAY-1998.  
 XX  
 XX 12-FEB-1997; 97US-0798691.  
 PE 12-FEB-1996; 96US-0598591.  
 PR 12-FEB-1997; 97US-0798691.  
 XX (ONCO-) ONCORMED INC.  
 PA Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
 PI Scheller DB, Zeng B;  
 PI WPI: 1998-296774/26.  
 DR BRCAl om1 gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 PS Claim 2e; Page -: 54pp; English.  
 XX This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) om13 gene in which a polymorphic variation occurs at  
 CC nucleotide 4427. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 om13 gene represented in AAV46450.  
 CC  
 CC Sequence 5711 BP; 1953 A; 1098 C; 1277 G; 1382 T; 1 other;  
 SQ  
 Query Match 100.0%; Score 201; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGCACAGCAGAAACCTACACTCATGGAAGGTAAAGACCTGCACTGAGCCAAAG 60  
 Db 2100 AGGCACAGCAGAAACCTACACTCATGGAAGGTAAAGACCTGCACTGAGCCAAAG 2159  
 QY 61 AGTACACAGCCAAATGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 120  
 Db 2160 AGTACACAGCCAAATGACAGACAGTAAGACATGATGATCTTCCAGAGCTG 2219  
 QY 121 AAGTTAACAATGACACCTGCTCTTTACTAAGTGTCAATGCACTGAAGTTAAAGAA 180  
 Db 2220 AAGTTAACAATGACACCTGCTCTTTACTAAGTGTCAATGCACTGAAGTTAAAGAA 2279  
 QY 181 TTGTCAATCTAGCCTTCCA 201  
 Db 2280 TTGTCAATCTAGCCTTCCA 2300  
 RESULT 8  
 AAV46471

ID	AAV46471 standard; cDNA: 5711 BP	
XX	AAV46471;	
XX	18-NOV-1998 (first entry)	
XX	Human BRCA1 omi3 polymorphism #7 cDNA.	
XX	BRCA1; omi1; human; breast and ovarian cancer predisposing gene;	
KW	polymorphism; susceptibility; anti-oncogene; tumour suppressor;	
KW	chromosome 17q; ss.	
XX	Homo sapiens.	
OS		
FH	Key	Location/Qualifiers
FT	CDS	120..5711
FT		/*tag= a
FT	variation	/product= "BRCA1 omi3 protein"
FT		4956
FT	/*tag= b	
FT	/note= "This polymorphic variation can be an A or G	nucleotide"
XX		
XX	US5750400-A.	
XX	12-MAY-1998.	
XX	12-FEB-1997; 97US-0798691.	
XX	12-FEB-1996; 96US-0598591.	
PR	12-FEB-1997; 97US-0798691.	
XX	(ONCO-) ONCORMED INC.	
PA		
XX	Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;	
PI	Schelter DB, Zeng B;	
PI	WPI: 1998-296774/26.	
DR		
XX	BRCA1 omi gene coding sequences - useful for distinguishing between	
PT	polymorphisms and mutation(s) in the screening for disposition to	
PT	breast or ovarian cancer	
XX		
PS	Claim 2e; Page -: 54pp; English.	
XX		
CC	This sequence encodes a human BRCA1 (breast and ovarian cancer	
CC	predisposing gene) omi3 gene in which a polymorphic variation occurs at	
CC	nucleotide 4956. This sequence and other polymorphic variations of this	
CC	sequence are useful for the identification of an individual who may or	
CC	may not have an increased susceptibility to breast or ovarian cancer.	
CC	The sequences used identify gene changes which are due to polymorphisms,	
CC	rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour	
CC	suppressor) which is involved in genetic inheritance of cancers,	
CC	especially breast and ovarian cancer. It is found at human chromosome	
CC	17q which is known to be linked to cancer susceptibility, especially	
CC	breast cancer. Cells containing a mutation in this gene lose the	
CC	wild-type function of BRCA1 and are more susceptible to cancers.	
CC	NOTE: This sequence does not appear in the specification but has been	
CC	created from the wild type BRCA1 omi3 gene represented in AAV46450.	
XX		
XX	Sequence 5711 BP; 1953 A; 1098 C; 1276 G; 1383 T; 1 other;	
SQ		
	Query Match	100.0%; Score 201; DB 19; Length 5711;
	Best Local Similarity	100.0%; Pred. No. 6,4e-97;
	Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 AGGCACAGCAAACTCTCAACTGTGAAGTAAGAACTGCACTGGAGCCAGAAG 60	
DB	2100 AGGCACAGCAAACTCTCAACTGTGAAGTAAGAACTGCACTGGAGCCAGAAG 2159	
	61 AGTAAAGCAAGCAAAATGAACAGCAAAAGTAAAGACATGACACTGTACTTCCAGAGCTG 120	
	2160 AGTAAAGCAAGCAAAATGAACAGCAAAAGTAAAGACATGACACTGTACTTCCAGAGCTG 2219	

```

OY      121  AAGTTAAACAATGCACCTGGTCTTTCTTACTAGTCTCAATACCACTGAATTTAAAGAA 180
Db      2220  AAGTTAAACAATGCACCTGGTCTTTCTTACTAGTCTCAATACCACTGAATTTAAAGAA 2277
OY      181  TTTGTCAATCTTACCTTCCA 201
Db      2280  TTTGTCAATCTTACCTTCCA 2300

RESULT 9
AAV46450 ID AAV46450 standard; cDNA; 5711 bp.
AC AAV46450;
XX
XX 18-NOV-1998 (first entry)
XX
XX Human BRCA1 omi3 cDNA.
XX
XX BRCA1: omi3: human; breast and ovarian cancer predisposing gene;
KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;
KM chromosome 17q: ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 120..5711
FT CDS /*tag= a
FT /product= "BRCA1 omi3 protein"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX 12-FEB-1996; 96US-0598591.
XX 12-FEB-1997; 97US-0798691.
XX
XX (ONCO-) ONCORMED INC.
XX
XX Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
XX Schelter DB, Zeng B;
XX
XX WPI: 1998-296774/26.
XX P-PSDB: AAW76100.
XX
XX BRCA1 omi gene coding sequences - useful for distinguishing between
XX polymorphisms and mutation(s) in the screening for disposition to
XX breast or ovarian cancer
XX
XX Claim 2e; Column 55-62; 54pp; English.
XX
XX This sequence encodes the human BRCA1 (breast and ovarian cancer
XX predisposing gene) omi3 gene. This sequence and polymorphic variations of
XX this sequence are useful for the identification of an individual who may
XX or may not have an increased susceptibility to breast or ovarian cancer.
XX The sequences used identify gene changes which are due to polymorphisms,
XX rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
XX suppressor) which is involved in genetic inheritance of cancers,
XX especially breast and ovarian cancer. It is found at human chromosome 17q
XX which is known to be linked to cancer susceptibility, especially breast
XX cancer. Cells containing a mutation in this gene lose the wild-type
XX function of BRCA1, and are more susceptible to cancers.
XX
XX Sequence 5711 bp; 1953 A; 1098 C; 1277 G; 1383 T; 0 other;
SQ
Query Match 100.0%; Score 201; DB 19; Length 5711;
Best Local Similarity 100.0%; Pred. No. 6 4e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AGGCACAGCAGAACTCAACTCATGGAAGCTAAAGAACCTTCGACCTGACCAAGAAG 60

```

```
Db 2100 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGAACTGCACTGAGCCAGAGAG 2159
Qy 61 AGTACACAGCCCAATGACAGACAGATAAAGACATGACATGACTTCTTCCAGAGCTG 120
Db 2160 AGTACACAGCCCAATGACAGACAGATAAAGACATGACATGACTTCTTCCAGAGCTG 2219
Qy 121 AAGTTACCAATGACAGCTGTTCTTTACTAGTGTCAATACCAATGACAGTAAAGAA 180
Db 2220 AAGTTACCAATGACAGCTGTTCTTTACTAGTGTCAATACCAATGACAGTAAAGAA 2279
Qy 181 TTGTCAATCCTACGCTTCCA 201
Db 2280 TTGTCAATCCTACGCTTCCA 2300

RESULT 10
AAV46452 standard; cDNA; 5711 BP.
AAV46452;
AC AAV46452;
XX 18-NOV-1998 (first entry)
DT
XX Human BRCA1 om1 polymorphism #2 cDNA.
DE
XX BRCA1: om1; human: breast and ovarian cancer predisposing gene:
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
KM chromosome 17q; ss.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT /tag= a
FT /product= "BRCA1 om1 protein"
FT variation 2430
FT /tag= b
FT /note= "This polymorphic variation can be a T or C
FT nucleotide"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX 12-FEB-1996; 96US-0598591.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX (ONCO-) ONCORMED INC.
XX
XX Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
XX Schelter DB, Zeng B;
XX
XX WPI: 1998-296774/26.
XX
XX BRCA1 om1 gene coding sequences - useful for distinguishing between
XX polymorphisms and mutation(s) in the screening for disposition to
XX breast or ovarian cancer
XX
XX Claim 2e; Page -: 54pp; English.
XX
XX This sequence encodes a human BRCA1 (breast and ovarian cancer
XX predisposing gene) om1 gene in which a polymorphic variation occurs at
XX nucleotide 2430. This sequence and other polymorphic variations of this
XX sequence are useful for the identification of an individual who may or
XX may not have an increased susceptibility to breast or ovarian cancer.
XX The sequences used identify gene changes which are due to polymorphisms,
XX rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
XX suppressor) which is involved in genetic inheritance of cancers,
XX especially breast and ovarian cancer. It is found at human chromosome
XX 17q which is known to be linked to cancer susceptibility, especially
```

```
CC breast cancer. Cells containing a mutation in this gene lose the
CC wild-type function of BRCA1 and are more susceptible to cancers.
CC NOTE: This sequence does not appear in the specification but has been
CC created from the wild type BRCA1 om1 gene represented in AAV46448.
XX
XX Sequence 5711 BP; 1953 A; 1098 C; 1277 G; 1382 T; 1 other;
S0

Query Match 100.0%; Score 201; DB 19; Length 5711;
Best Local Similarity 100.0%; Pred. NO. 6.4e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGAACTGCACTGAGCCAGAGAG 60
Db 2100 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGAACTGCACTGAGCCAGAGAG 2159
Qy 61 AGTACACAGCCCAATGACAGACAGATAAAGACATGACATGACTTCTTCCAGAGCTG 120
Db 2160 AGTACACAGCCCAATGACAGACAGATAAAGACATGACATGACTTCTTCCAGAGCTG 2219
Qy 121 AAGTTACCAATGACAGCTGTTCTTTACTAGTGTCAATACCAATGACAGTAAAGAA 180
Db 2220 AAGTTACCAATGACAGCTGTTCTTTACTAGTGTCAATACCAATGACAGTAAAGAA 2279
Qy 181 TTGTCAATCCTACGCTTCCA 201
Db 2280 TTGTCAATCCTACGCTTCCA 2300

RESULT 11
AAV46453 standard; cDNA; 5711 BP.
AAV46453;
AC AAV46453;
XX 18-NOV-1998 (first entry)
DT
XX Human BRCA1 om1 polymorphism #3 cDNA.
DE
XX BRCA1: om1; human: breast and ovarian cancer predisposing gene:
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
KM chromosome 17q; ss.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT /tag= a
FT /product= "BRCA1 om1 protein"
FT variation 2731
FT /tag= b
FT /note= "This polymorphic variation can be a C or T
FT nucleotide"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX 12-FEB-1996; 96US-0598591.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX (ONCO-) ONCORMED INC.
XX
XX Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
XX Schelter DB, Zeng B;
XX
XX WPI: 1998-296774/26.
XX
XX BRCA1 om1 gene coding sequences - useful for distinguishing between
XX polymorphisms and mutation(s) in the screening for disposition to
XX breast or ovarian cancer
```



PS Claim 2e; Page -: 54pp; English.  
XX This sequence encodes a human BRCA1 (breast and ovarian cancer  
CC predisposing gene) om1 gene in which a polymorphic variation occurs at  
CC nucleotide 2731. This sequence and other polymorphic variations of this  
CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers, especially  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 om1 gene represented in AAV46448.  
XX  
SQ Sequence 5711 BP; 1953 A; 1099 C; 1277 G; 1381 T; 1 other;  
Query Match 100.0%; Score 201; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGCACAGCAGAAACCTACATCATGAGGTAAGAACCTCGACCTGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAAACCTACATCATGAGGTAAGAACCTCGACCTGAGCCAGAG 2159  
QY 61 AGTAACAGCCCAATGATGACAGAGTAAGAGATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTAACAGCCCAATGATGACAGAGTAAGAGATGATGATCTTCCAGAGCTG 2219  
QY 121 AAGTTAACAAATGACACCTGCTTCTTTACTAAGTGTTCAAATACAGTGAATTAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTTCTTTACTAAGTGTTCAAATACAGTGAATTAAGAA 2279  
QY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2280 TTTGTCAATCCTAGCCTTCCA 2300  
RESULT 12  
AAV46454  
ID AAV46454 standard; cDNA; 5711 BP.  
XX  
AC AAV46454;  
XX  
DT 18-NOV-1998 (first entry)  
XX  
DE Human BRCA1 om1 polymorphism #4 cDNA.  
XX  
KM BRCA1; om1; human; breast and ovarian cancer predisposing gene;  
KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 120..5711  
FT FT /tag= a  
FT FT /product= "BRCA1 om1 protein"  
FT FT 3232  
FT FT /tag= b  
FT FT /note= "This polymorphic variation can be an A or G  
XX nucleotide"  
XX  
PM US5750400-A.  
XX  
PD 12-MAY-1998.  
XX  
PF 12-FEB-1997; 97US-0798691.  
XX  
PR 12-FEB-1996; 96US-0598591.  
PR 12-FEB-1997; 97US-0798691.

XX (ONCO-) ONCOMED INC.  
PA  
XX  
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Scheller DB, Zeng B;  
XX  
DR WPI: 1998-296774/26.  
XX  
PT BRCA1 om1 gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer  
XX  
PS Claim 2e; Page -: 54pp; English.  
XX  
CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
CC predisposing gene) om1 gene in which a polymorphic variation occurs at  
CC nucleotide 3232. This sequence and other polymorphic variations of this  
CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers, especially  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 om1 gene represented in AAV46448.  
XX  
SQ Sequence 5711 BP; 1953 A; 1099 C; 1276 G; 1382 T; 1 other;  
Query Match 100.0%; Score 201; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 6.4e-97;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGCACAGCAGAAACCTACATCATGAGGTAAGAACCTCGACCTGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAAACCTACATCATGAGGTAAGAACCTCGACCTGAGCCAGAG 2159  
QY 61 AGTAACAGCCCAATGATGACAGAGTAAGAGATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTAACAGCCCAATGATGACAGAGTAAGAGATGATGATCTTCCAGAGCTG 2219  
QY 121 AAGTTAACAAATGACACCTGCTTCTTTACTAAGTGTTCAAATACAGTGAATTAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTTCTTTACTAAGTGTTCAAATACAGTGAATTAAGAA 2279  
QY 181 TTTGTCAATCCTAGCCTTCCA 201  
DB 2280 TTTGTCAATCCTAGCCTTCCA 2300  
RESULT 13  
AAV46455  
ID AAV46455 standard; cDNA; 5711 BP.  
XX  
AC AAV46455;  
XX  
DT 18-NOV-1998 (first entry)  
XX  
DE Human BRCA1 om1 polymorphism #5 cDNA.  
XX  
KM BRCA1; om1; human; breast and ovarian cancer predisposing gene;  
KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 120..5711  
FT FT /tag= a  
FT FT /product= "BRCA1 om1 protein"  
FT FT 3667  
FT FT variation

```
FT FT /*tag= b
FT /note= "This polymorphic variation can be an A or G
XX nucleotide"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX 12-FEB-1996; 96US-0598591.
XX 12-FEB-1997; 97US-0798691.
XX (ONCO-) ONCORMED INC.
XX
XX Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
XX Schelter DB, Zeng B;
XX WPI; 1998-296774/26.
XX
XX BRCA1 omi gene coding sequences - useful for distinguishing between
XX PT polymorphisms and mutation(s) in the screening for disposition to
XX PT breast or ovarian cancer
XX
XX Claim 2e; Page -: 54pp; English.
XX
XX This sequence encodes a human BRCA1 (breast and ovarian cancer
XX CC predisposing gene) omi1 gene in which a polymorphic variation occurs at
XX CC nucleotide 3667. This sequence and other polymorphic variations of this
XX CC sequence are useful for the identification of an individual who may or
XX CC may not have an increased susceptibility to breast or ovarian cancer.
XX CC The sequences used identify gene changes which are due to polymorphisms,
XX CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
XX CC suppressor) which is involved in genetic inheritance of cancers,
XX CC especially breast and ovarian cancer. It is found at human chromosome
XX CC 17q which is known to be linked to cancer susceptibility, especially
XX CC breast cancer. Cells containing a mutation in this gene lose the
XX CC wild-type function of BRCA1 and are more susceptible to cancers.
XX CC NOTE: This sequence does not appear in the specification but has been
XX CC created from the wild type BRCA1 omi1 gene represented in AAV6448.
XX
XX Sequence 5711 BP; 1953 A; 1099 C; 1276 G; 1382 T; 1 other:
SQ
Query Match 100.0%; Score 201; DB 19; Length 5711;
Best Local Similarity 100.0%; Pred. No. 6.4e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGGCACAGCAGAAACCTACATCTATGGAAGTAAGAACCTGAGCCAGAG 60
DB 2100 AGGCACAGCAGAAACCTACATCTATGGAAGTAAGAACCTGAGCCAGAG 2159
QY 61 AGTAACAGGCAATGATGACAGACATGAAGACATGATGATCTTCCAGAGCTG 120
DB 2160 AGTAACAGGCAATGATGACAGACATGAAGACATGATGATCTTCCAGAGCTG 2219
QY 121 AAGTTACAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATCCAGTAACTTAAGAA 180
DB 2220 AAGTTACAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATCCAGTAACTTAAGAA 2279
QY 181 TTTGTCAATCTTAGCCTTCCA 201
DB 2280 TTTGTCAATCTTAGCCTTCCA 2300
RESULT 14
AAV6448
ID AAV6448 standard; cDNA; 5711 BP.
XX
XX AAV6448;
AC
XX
XX 18-NOV-1998 (first entry)
DT
XX
XX Human BRCA1 omi1 polymorphism #6 cDNA.
DE
```

```
XX
XX BRCA1; omi1; human; breast and ovarian cancer predisposing gene;
XX KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
XX KW chromosome 17q; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 120..5711
XX FT CDS
XX FT
XX FT variation
XX FT /tag= b
XX FT /note= "This polymorphic variation can be a T or C
XX FT nucleotide"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 97US-0798691.
XX
XX 12-FEB-1996; 96US-0598591.
XX 12-FEB-1997; 97US-0798691.
XX (ONCO-) ONCORMED INC.
XX
XX Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
XX PI Schelter DB, Zeng B;
XX WPI; 1998-296774/26.
XX
XX BRCA1 omi gene coding sequences - useful for distinguishing between
XX PT polymorphisms and mutation(s) in the screening for disposition to
XX PT breast or ovarian cancer
XX
XX Claim 2e; Page -: 54pp; English.
XX
XX This sequence encodes a human BRCA1 (breast and ovarian cancer
XX CC predisposing gene) omi1 gene in which a polymorphic variation occurs at
XX CC nucleotide 4427. This sequence and other polymorphic variations of this
XX CC sequence are useful for the identification of an individual who may or
XX CC may not have an increased susceptibility to breast or ovarian cancer.
XX CC The sequences used identify gene changes which are due to polymorphisms,
XX CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
XX CC suppressor) which is involved in genetic inheritance of cancers,
XX CC especially breast and ovarian cancer. It is found at human chromosome
XX CC 17q which is known to be linked to cancer susceptibility, especially
XX CC breast cancer. Cells containing a mutation in this gene lose the
XX CC wild-type function of BRCA1 and are more susceptible to cancers.
XX CC NOTE: This sequence does not appear in the specification but has been
XX CC created from the wild type BRCA1 omi1 gene represented in AAV6448.
XX
XX Sequence 5711 BP; 1953 A; 1098 C; 1277 G; 1382 T; 1 other:
SQ
Query Match 100.0%; Score 201; DB 19; Length 5711;
Best Local Similarity 100.0%; Pred. No. 6.4e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGGCACAGCAGAAACCTACATCTATGGAAGTAAGAACCTGAGCCAGAG 60
DB 2100 AGGCACAGCAGAAACCTACATCTATGGAAGTAAGAACCTGAGCCAGAG 2159
QY 61 AGTAACAGGCAATGATGACAGACATGAAGACATGATGATCTTCCAGAGCTG 120
DB 2160 AGTAACAGGCAATGATGACAGACATGAAGACATGATGATCTTCCAGAGCTG 2219
QY 121 AAGTTACAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATCCAGTAACTTAAGAA 180
DB 2220 AAGTTACAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATCCAGTAACTTAAGAA 2279
QY 181 TTTGTCAATCTTAGCCTTCCA 201
DB 2280 TTTGTCAATCTTAGCCTTCCA 2300
```

Db 2280 TTTGTCATCCTAGCCTTCCA 2300

RESULT 15  
DAVAGEAC7

ID	AAV46457	standard; cDNA; 5711 BP.
1	AAV46457	standard; cDNA; 5711 BP.

AAV46457;  
AC

DT 18-NOV-1998 (first entry)  
XX

Human BRCA1 om1 polymorphism #7 cDNA

KW BRCA1; omil; human; breast and ovarian cancer predisposing gene  
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	120. .5711

120..5711

```
/*tag= a
/product= "BPCL oil" protect="
```

FT variation

FT /note= "This polymorphic variation can be an A or G  
FT nucleotide"

PN US5750400-A.

PD 12-MAY-1998

PF 12-FEB-1997; 97US-0798691.

PR 12-FEB-1996; 96US-0598591.

PR 12-FEB-1997; 97US-0798691.  
YY

PA (ONCO-) ONCORMED INC.  
XX

PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Scholten DR, Zeng H.

PI Schelter DB, Zeng B, XX

DR WPI; 1.998-296774/26.

PT BRCA1 omi gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer

PS Claim 2e; Page -; 54pp; English.

CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
CC predisposing gene) om1 gene in which a polymorphic variation occurs at  
CC nucleotide 4956. This sequence and other polymorphic variations of this  
CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms  
CC rather than cancer involving mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers,  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 om1 gene represented in AAV6448.

SQ Sequence 5711..BP; 1953 A; 1099 C; 1276 G; 1382 T; 1 other,

Query Match	100.0%	Score 201;	DB 19;	Length 5711;
-------------	--------	------------	--------	--------------

Best Local Similarity 100.0%; Pred. No. 6.4e-97;

Matches	201;	Conservative	0;	Mismatches	0;	Indels	0;	. Gaps	0;
---------	------	--------------	----	------------	----	--------	----	--------	----

```

Db      2100  AAGCCAGCAGCAAAACCTCAACTCATGTGGAAGAGTAAAGAACTGCMAACTGAGACCAAGAG 2159
QY      61    AGTAACAGCCCAATGTGAACACAGCAAGTAAAGACACTGACAGTACTTTCCCAAGAGCTG 120
        |||||
Db      2160  AGTAAACAAGCCAAATGAAGACACAGTAAAGACACTGACAGTGAATCTTTCCCAAGAGCTG 2219
QY      121  AAGTTAACAAATGCACTCGTCTTTCTACTAAGTTGTTCAAAATACCAGAGCACTTAAAGAA 180
        |||||
Db      2220  AAGTTAACAAATGCACTCGTCTTTCTTTCTAAGTGTTCAAATACCAGAGCACTTAAAGAA 2279
QY      181  TTTGTCAATCCTAGCCTTCCA 201
        |||||
Db      2280  TTTGTCAATCCTAGCCTTCCA 2300

```

61 AGTACCAAGCCAAATGAACAGACAAGTAAAAGACATGACAGTGTACTTTCCAGAGCTG 120

Db 2160 AGTAACAAGCCAAATGAACAGACAGTAAAGAATGACAGTGTACTTTCCAGAGCTG 2219

QY 121 AAGTTACCAATGCACCTGGTCTTTACTAAGTGTTCAAATACCAAGTGAACCTTAAGAA 180

Db 2220 AAGTTAACAAATGCACCTGGTTCCTTTTACTAAGTGTTCAAATACCAGTGAACCTTAAGAA 2279

181 TTGTCATCCTAGCCTTCCA 201

Db 2280 TTGTCAATCCTAGCCTTCCA 2300

```
Search completed: June 27, 2003, 05:28:56
Job time : 248 secs
```

Job time : 248 secs

Fri Jun 27 10:47:45 2003

us-09-734-672-3\_copy\_2100\_2300.011.mpb

C to T at 22201

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 01:02:23 ; Search time 149 Seconds  
(Without alignments)  
2002.511 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_2100\_2300  
Perfect score: 201  
Sequence: 1 AGGCACAGCAGAACTACA.....TTGTCAATCCTACCTTCCA 201

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 1055720 seqs, 742224136 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	201	100.0	5711	9 US-09-734-672-1	Sequence 1, Appl1
2	201	100.0	5711	9 US-09-734-672-5	Sequence 5, Appl1
3	201	100.0	5711	9 US-09-734-672-1	Sequence 1, Appl1
4	201	100.0	5711	9 US-09-882-828-1	Sequence 3, Appl1
5	201	100.0	5711	9 US-09-882-828-3	Sequence 1, Appl1
6	150	74.6	5711	9 US-09-734-672-3	Sequence 3, Appl1
7	150	74.6	5711	9 US-09-882-828-5	Sequence 5, Appl1
8	84	41.8	121	9 US-09-818-875-765	Sequence 765, App
9	84	41.8	121	9 US-09-818-875-766	Sequence 761, App
10	70	34.8	121	9 US-09-818-875-761	Sequence 762, App
11	70	34.8	121	9 US-09-818-875-762	Sequence 8, Appl1
12	31	15.4	31	10 US-09-838-497-8	Sequence 3, Appl1
13	24	11.4	24	9 US-09-734-672-3	Sequence 32, Appl1
14	23	11.4	23	9 US-09-882-828-3	Sequence 28, Appl1
15	23	11.4	23	9 US-10-022-819-28	Sequence 96, Appl1
16	23	11.4	23	9 US-09-823-327-96	Sequence 1, Appl1
17	20	10.0	24	9 US-10-259-479-1	Sequence 2378, Ap
18	19	9.5	282	10 US-09-294-093B-2378	

20	18	9.0	293	10 US-09-867-701-339	Sequence 339, App
21	18	9.0	309	10 US-09-867-701-410	Sequence 410, App
22	18	9.0	505	10 US-09-864-761-13432	Sequence 13432, A
23	18	9.0	1763	10 US-09-917-800A-1460	Sequence 1460, Ap
24	17	8.5	17	9 US-09-818-875-767	Sequence 767, App
25	17	8.5	17	9 US-09-818-875-768	Sequence 768, App
26	17	8.5	17	9 US-09-923-327-170	Sequence 170, App
27	17	8.5	17	9 US-09-923-327-189	Sequence 189, App
28	17	8.5	237	9 US-10-054-387-74	Sequence 74, Appl1
29	17	8.5	284	10 US-09-878-574-8061	Sequence 8061, Ap
30	17	8.5	648	9 US-09-991-936-1599	Sequence 1599, Ap
31	17	8.5	4842	9 US-10-174-590-289	Sequence 289, App
32	17	8.5	4842	9 US-10-176-758-289	Sequence 289, App
33	17	8.5	4842	9 US-10-175-737-289	Sequence 289, App
34	17	8.5	4842	9 US-10-173-706-289	Sequence 289, App
35	17	8.5	4842	9 US-10-175-738-289	Sequence 289, App
36	17	8.5	4842	9 US-10-175-752-289	Sequence 289, App
37	17	8.5	4842	9 US-10-176-482-289	Sequence 289, App
38	17	8.5	4842	9 US-10-176-757-289	Sequence 289, App
39	17	8.5	4842	9 US-10-176-913-289	Sequence 289, App
40	17	8.5	4842	9 US-10-180-557-289	Sequence 289, App
41	17	8.5	4842	9 US-10-180-557-289	Sequence 289, App
42	17	8.5	4842	9 US-10-173-700-289	Sequence 289, App
43	17	8.5	4842	9 US-10-174-572-289	Sequence 289, App
44	17	8.5	4842	9 US-10-174-579-289	Sequence 289, App
45	17	8.5	4842	9 US-10-174-582-289	Sequence 289, App

#### ALIGNMENTS

RESULT 1  
US-09-734-672-1  
Sequence 1, Application US/09734672  
Publication No. US20020183268A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvarez, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
BRCA1 Gene  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESS: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
City: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAl  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-734-672-1

Query Match 100.0%; Score 201; DB 9; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 8.9e-100;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTACACTCATGAGGTAAAGCACTGAGGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAACTACACTCATGAGGTAAAGCACTGAGGAGCCAGAG 2159  
QY 61 AGTAACAAGCCAAATGAGACAGAGTAAGACATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTAACAAGCCAAATGAGACAGAGTAAGACATGATGATCTTCCAGAGCTG 2219  
QY 121 AAGTTAACAAATGACACCTGCTTTTACTAAGTGTCAATACGAGTAAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTTTTACTAAGTGTCAATACGAGTAAAGAA 2279  
QY 181 TTGTCAATCTAGCCTTCCA 201  
DB 2280 TTGTCAATCTAGCCTTCCA 2300

RESULT 2  
US-09-734-672-5  
Sequence 5, Application US/09734672  
Publication No. US20020183268A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
BRCAl Gene  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/596,591  
FILING DATE: 12-Feb-96  
ATTORNEY/AGENT INFORMATION:

NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAl  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-734-672-5

Query Match 100.0%; Score 201; DB 9; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 8.9e-100;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTACACTCATGAGGTAAAGCACTGAGGAGCCAGAG 60  
DB 2100 AGGCACAGCAGAACTACACTCATGAGGTAAAGCACTGAGGAGCCAGAG 2159  
QY 61 AGTAACAAGCCAAATGAGACAGAGTAAGACATGATGATCTTCCAGAGCTG 120  
DB 2160 AGTAACAAGCCAAATGAGACAGAGTAAGACATGATGATCTTCCAGAGCTG 2219  
QY 121 AAGTTAACAAATGACACCTGCTTTTACTAAGTGTCAATACGAGTAAAGAA 180  
DB 2220 AAGTTAACAAATGACACCTGCTTTTACTAAGTGTCAATACGAGTAAAGAA 2279  
QY 181 TTGTCAATCTAGCCTTCCA 201  
DB 2280 TTGTCAATCTAGCCTTCCA 2300

RESULT 3  
US-09-982-828-1  
Sequence 1, Application US/09982828  
Publication No. US20030022184A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Thurber, Denise  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
BRCAl Gene  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Avenue N. W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/982,828  
FILING DATE: 22-Oct-2001

```

      ZIP: 20004
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/982,828
        FILING DATE: 22-Oct-2001
        CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 09/074,453
        FILING DATE: 1998-05-06
        APPLICATION NUMBER: US 08/798,691
        FILING DATE: 1997-02-12
        APPLICATION NUMBER: US 08/598,591
        FILING DATE: 1996-02-12
      ATTORNEY/AGENT INFORMATION:
        NAME: Michael S. Tuscan
        REGISTRATION NUMBER: 43,210
        REFERENCE/DOCKET NUMBER: 44921-5053-01-US
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 202-739-3000
        TELEFAX: 202-739-3001
      INFORMATION FOR SEQ ID NO: 3:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 5711 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: not relevant
          TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      ORIGINAL SOURCE:
        ORGANISM: Homo sapiens
        STRAIN: BRCAL (om12)
        POSITION IN GENOME:
          CHROMOSOME/SEGMENT: 17
          MAP POSITION: 17q21
      SEQUENCE DESCRIPTION: SEQ ID NO: 3:
      US-09-982-828-3

Query Match           100.0%; Score 201; DB 9; Length 5711;
Best Local Similarity 100.0%; Pred.No.8,9e-100;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   1 AGGCACAGCAGAAAACCTACACTGATGAGAGGTAAAGAACCCTGCACAACCTGAGCCCAAGAG 60
Db   2100 AGGCACAGCAGAAAACCTACACTGATGAGAGGTAAAGAACCCTGCACAACCTGAGCCCAAGAG 2159
QY   .61 AGTAACAAGCCAATTGAACACACAAAGTAAAGACATGACATGATCTTTCCACAGCTG 120
Db   2160 AGTAACAAGCCAATTGAACACACAAAGTAAAGACATGACATGATCTTTCCACAGCTG 2219
QY   121 AAGTTAACAAATGACACCTGGTTCTTTTACTAAGTGTTCAAATPACACAGTAGCACTTAAGAA 180
Db   2220 AAGTTAACAAATGACACCTGGTTCTTTTACTAAGTGTTCAAATPACACAGTAGCACTTAAGAA 2279
QY   181 TTGTGCAATCCTAGCCTTCCA 201
Db   2280 TTGTGCAATCCTAGCCTTCCA 2300

RESULT 5
US-10-022-819-1
Sequence 1, Application US/10022819
Publication No. US20030027166A1
GENERAL INFORMATION:
APPLICANT: ALLEN, Antoinette C. P.
OLSEN, Sheri J.
LAWRENCE, Tammy
ANGELLY, Tracy S.
RABIN, Mark B.
TITLE OF INVENTION: CODING HAPLOTYPE OF THE HUMAN
BRCAL GENE

```

NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Avenue  
CITY: Washington DC  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/022,819  
FILING DATE: 22-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/074,452  
FILING DATE: 1998-05-06  
ATTORNEY/AGENT INFORMATION:  
NAME: <Unknown>  
REGISTRATION NUMBER: <Unknown>  
REFERENCE/DOCKET NUMBER: 044921-5049-01-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3001  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRADEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: HOMO SAPIENS  
STRAIN: BRCAL  
HAPLOTYPE: OM14  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-022-819-1  
Query Match 100.0%; Score 201; DB 9; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 8.9e-100;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTACACTGAGAGTAAGAAACCTGCACACTGGAGCCAGAG 60  
|||||  
DB 2100 AGGCACAGCAGAACTACACTGAGAGTAAGAAACCTGCACACTGGAGCCAGAG 2159  
61 AGTAACAAGCCAAATGACAGACAGTAAGAAAGCATGACAGTATCTTCCAGAGCTG 120  
|||||  
DB 2160 AGTAACAAGCCAAATGACAGACAGTAAGAAAGCATGACAGTATCTTCCAGAGCTG 2219  
QY 121 AGTTAACAAGTACGACCTGTTCTTTACTAAGTGTTCAAATACCAAGTAAAGAA 180  
|||||  
DB 2220 AAGTTAACAAGTACGACCTGTTCTTTACTAAGTGTTCAAATACCAAGTAAAGAA 2279  
QY 181 TTTGTCAATCTAGCCTTCCA 201  
|||||  
DB 2280 TTTGTCAATCTAGCCTTCCA 2300

RESULT 6  
US-09-734-672-3  
Sequence 3, Application US/09734672  
Publication No. US20020183268A1  
GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Scheller, Denise B.  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
BRCAL Gene  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRADEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAL  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-734-672-3  
Query Match 74.6%; Score 150; DB 9; Length 5711;  
Best Local Similarity 99.5%; Pred. No. 6.4e-72;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTACACTGAGAGTAAGAAACCTGCACACTGGAGCCAGAG 60  
|||||  
DB 2100 AGGCACAGCAGAACTACACTGAGAGTAAGAAACCTGCACACTGGAGCCAGAG 2159  
61 AGTAACAAGCCAAATGACAGACAGTAAGAAAGCATGACAGTATCTTCCAGAGCTG 120  
|||||  
DB 2160 AGTAACAAGCCAAATGACAGACAGTAAGAAAGCATGACAGTATCTTCCAGAGCTG 2219  
QY 121 AAGTTAACAAGTACGACCTGTTCTTTACTAAGTGTTCAAATACCAAGTAAAGAA 180  
|||||  
DB 2220 AAGTTAACAAGTACGACCTGTTCTTTACTAAGTGTTCAAATACCAAGTAAAGAA 2279  
QY 181 TTTGTCAATCTAGCCTTCCA 201  
|||||  
DB 2280 TTTGTCAATCTAGCCTTCCA 2300



RESULT 7  
US-09-828-828-5  
Sequence 5, Application US/09982828  
Publication No. US20030022184A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Thurber, Denise  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
BRCA1 Gene  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Avenue N. W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/982,828  
FILING DATE: 22-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 09/074,453  
FILING DATE: 1998-05-06  
APPLICATION NUMBER: US 08/798,691  
FILING DATE: 1997-02-12  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 1996-02-12  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCA1 (om13)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-828-828-5  
Query Match 74.6% Score 150; DB 9; Length 5711;  
Best Local Similarity 99.5% Pred. No. 6.4e-72;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCACAGCAGAACTACATGAGAGTAAGAACTGCACTGGAGCAGAG 60  
|||||  
DB 2100 AGGCACAGCAGAACTACATGAGAGTAAGAACTGCACTGGAGCAGAG 2159  
QY 61 AGTACACAGCAGAACTACATGAGAGTAAGAACTGCACTGGAGCAGAG 120  
|||||  
DB 2160 AGTACACAGCAGAACTACATGAGAGTAAGAACTGCACTGGAGCAGAG 2219

QY 121 AAGTTACAAATGACACCTGGTCTTTTACTAGTGTCAATAACAGTGAACCTTAAGAA 180  
|||||  
DB 2220 AAGTTACAAATGACACCTGGTCTTTTACTAGTGTCAATAACAGTGAACCTTAAGAA 2279  
QY 181 TTTGTCAATCCTAGCCTTCCA 201  
|||||  
DB 2280 TTTGTCAATCCTAGCCTTCCA 2300

RESULT 8  
US-09-818-875-765  
Sequence 765, Application US/09818875  
Publication No. US20030051270A1  
GENERAL INFORMATION:  
APPLICANT: Kmiec, Eric B.  
Gamper, Howard B.  
APPLICANT: Rice, Michael C.  
TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
FILE REFERENCE: Napro-4  
CURRENT APPLICATION NUMBER: US/09/818,875  
CURRENT FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: US 60/192,176  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/192,179  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/208,538  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: US 60/244,989  
NUMBER OF SEQ ID NOS: 4385  
SOFTWARE: Friedman macro Napro4  
SEQ ID NO 765  
LENGTH: 121  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-818-875-765  
Query Match 41.8% Score 84; DB 9; Length 121;  
Best Local Similarity 100.0% Pred. No. 5.9e-36;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 CTGAGTTACAAATGACACCTGGTCTTTTACTAGTGTCAATAACAGTGAACCTTA 177  
|||||  
DB 1 CTGAGTTACAAATGACACCTGGTCTTTTACTAGTGTCAATAACAGTGAACCTTA 60  
QY 178 GAATTTGTCAATCCTAGCCTTCCA 201  
|||||  
DB 61 GAATTTGTCAATCCTAGCCTTCCA 84

RESULT 9  
US-09-818-875-766/c  
Sequence 766, Application US/09818875  
Publication No. US20030051270A1  
GENERAL INFORMATION:  
APPLICANT: Kmiec, Eric B.  
Gamper, Howard B.  
APPLICANT: Rice, Michael C.  
TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
FILE REFERENCE: Napro-4  
CURRENT APPLICATION NUMBER: US/09/818,875  
CURRENT FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: US 60/192,176  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/192,179  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/208,538  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: US 60/244,989  
PRIOR FILING DATE: 2000-10-30

NUMBER OF SEQ ID NOS: 4385  
 SOFTWARE: Friedman macro Napro4  
 SEQ ID NO 766  
 LENGTH: 121  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-818-875-766

Query Match  
 Best Local Similarity 41.8%; Score 84; DB 9; Length 121;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 CTGAAGTAAACAATGACCTGGTCTTTTACTAAGTGTCAATACCACTGAAGTAA 177  
 |||||||  
 Db 121 CTGAAGTAAACAATGACCTGGTCTTTTACTAAGTGTCAATACCACTGAAGTAA 62

QY 178 GAATTTGTCATCTAGCTTCCA 201  
 |||||||  
 Db 61 GAATTTGTCATCTAGCTTCCA 38

RESULT 10  
 US-09-818-875-761

Sequence 761, Application US/09818875  
 Publication No. US20030051270A1

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric B.

APPLICANT: Gamper, Howard B.

APPLICANT: Rice, Michael C.

TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
 FILE REFERENCE: Napro-4

CURRENT APPLICATION NUMBER: US/09/818,875

PRIOR FILING DATE: 2001-03-27

PRIOR APPLICATION NUMBER: US 60/192,176

PRIOR FILING DATE: 2000-03-27

PRIOR APPLICATION NUMBER: US 60/192,179

PRIOR FILING DATE: 2000-03-27

PRIOR APPLICATION NUMBER: US 60/208,538

PRIOR FILING DATE: 2000-06-01

PRIOR APPLICATION NUMBER: US 60/244,989

PRIOR FILING DATE: 2000-10-30

NUMBER OF SEQ ID NOS: 4385

SOFTWARE: Friedman macro Napro4

SEQ ID NO 761

LENGTH: 121

TYPE: DNA

ORGANISM: Homo sapiens

US-09-818-875-761

Query Match  
 Best Local Similarity 34.8%; Score 70; DB 9; Length 121;  
 Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 37 GAACCTGCACTGGAGCCAGAGTAACAGCCCAATGACAGACAGTAAGACAT 96  
 |||||||  
 Db 1 GAACCTGCACTGGAGCCAGAGTAACAGCCCAATGACAGACAGTAAGACAT 60

QY 97 GACAGTATACCTTCCAGAGCTGAAGTTAACAAATGACCGTCTTTTACTAAGTGT 156  
 |||||||  
 Db 61 GACAGTATACCTTCCAGAGCTGAAGTTAACAAATGACCGTCTTTTACTAAGTGT 120

QY 157 T 157  
 ||  
 Db 121 T 121

RESULT 11  
 US-09-818-875-762/C  
 Sequence 762, Application US/09818875  
 Publication No. US20030051270A1  
 GENERAL INFORMATION:  
 APPLICANT: Kmiec, Eric B.

APPLICANT: Gamper, Howard B.  
 APPLICANT: Rice, Michael C.  
 TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
 FILE REFERENCE: Napro-4

CURRENT APPLICATION NUMBER: US/09/818,875

PRIOR FILING DATE: 2001-03-27

PRIOR APPLICATION NUMBER: US 60/192,176

PRIOR FILING DATE: 2000-03-27

PRIOR APPLICATION NUMBER: US 60/192,179

PRIOR FILING DATE: 2000-03-27

PRIOR APPLICATION NUMBER: US 60/208,538

PRIOR FILING DATE: 2000-06-01

PRIOR APPLICATION NUMBER: US 60/244,989

PRIOR FILING DATE: 2000-10-30

NUMBER OF SEQ ID NOS: 4385

SOFTWARE: Friedman macro Napro4

SEQ ID NO 762

LENGTH: 121

TYPE: DNA

ORGANISM: Homo sapiens

US-09-818-875-762

Query Match  
 Best Local Similarity 34.8%; Score 70; DB 9; Length 121;  
 Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 37 GAACCTGCACTGGAGCCAGAGTAACAGCCCAATGACAGACAGTAAGACAT 96  
 |||||||  
 Db 121 GAACCTGCACTGGAGCCAGAGTAACAGCCCAATGACAGACAGTAAGACAT 62

QY 97 GACAGTATACCTTCCAGAGCTGAAGTTAACAAATGACCGTCTTTTACTAAGTGT 156  
 |||||||  
 Db 61 GACAGTATACCTTCCAGAGCTGAAGTTAACAAATGACCGTCTTTTACTAAGTGT 2

QY 157 T 157  
 ||  
 Db 1 T 1

RESULT 12  
 US-09-838-497-8

Sequence 8, Application US/09838497

Patent No. US20020061523A1

GENERAL INFORMATION:

APPLICANT: Schwartz, David C

APPLICANT: Wu, Tian

TITLE OF INVENTION: Method for Analyzing Nucleic Acid Reactions

FILE REFERENCE: 09820148

CURRENT APPLICATION NUMBER: US/09/838,497

PRIOR FILING DATE: 2001-04-19

PRIOR APPLICATION NUMBER: US 09/175,824

PRIOR FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 31

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: primer

US-09-838-497-8

Query Match  
 Best Local Similarity 15.4%; Score 31; DB 10; Length 31;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 CCAAGTGAAGTAAAGATTGTCAATCTAG 194  
 |||||||  
 Db 1 CCAAGTGAAGTAAAGATTGTCAATCTAG 31

RESULT 13

US-10-259-479-3  
; Sequence 3, Application US/10259479  
; Publication No. US20030082616A1  
; GENERAL INFORMATION:  
; APPLICANT: HITACH, LTD.  
; APPLICANT: National Cancer Center  
; TITLE OF INVENTION: Genetic Screening Method and Genetic Screening Apparatus  
; FILE REFERENCE: PH-1080  
; CURRENT APPLICATION NUMBER: US/10/259,479  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: JP 294257/1999  
; PRIOR FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 3  
; LENGTH: 24  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: DNA primer used for PCR.  
US-10-259-479-3

Query Match 11.9%; Score 24; DB 9; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.0032; Mismatches 0; Indels 0; Gaps 0;

QY 43 GCAACTGAGCCAGAGAGATAAC 66  
Db 1 GCACTGAGCCAGAGAGATAAC 24

## RESULT 14

US-09-734-672-32/C  
; Sequence 32, Application US/09734672  
; Publication No. US20020183268A1

## GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin

TITLE OF INVENTION: Coding Sequences of the Human  
BRCA1 Gene

NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96

## ATTORNEY/AGENT INFORMATION:

NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
STRAIN: 11ER primer  
SEQUENCE DESCRIPTION: SEQ ID NO: 32:  
US-09-734-672-32

Query Match 11.4%; Score 23; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.011; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGTTCAAATACGAGTAA 176  
Db 23 TGTTCAAATACGAGTAA 1

## RESULT 15

US-09-982-828-32/C  
; Sequence 32, Application US/09982828  
; Publication No. US20030022184A1

## GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Thurber, Denise  
Zeng, Bin

TITLE OF INVENTION: Coding Sequences of the Human  
BRCA1 Gene

NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Avenue N. W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/982,828  
FILING DATE: 22-Oct-2001  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/074,453  
FILING DATE: 1998-05-06  
APPLICATION NUMBER: US 08/798,691  
FILING DATE: 1997-02-12  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 1996-02-12

## ATTORNEY/AGENT INFORMATION:

NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:

STRAIN: 11ER primer  
 SEQUENCE DESCRIPTION: SEQ ID NO: 32:  
 US-09-982-828-32

Query Match 11.4%; Score 23; DB 9; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.011;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 TGTCAATACCAGTGAAGTAA 176  
 |||||  
 Db 23 TGTCAATACCAGTGAAGTAA 1

Search completed: June 27, 2003, 06:05:20  
 Job time : 159 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2003, 10:03:36 ; Search time 90 Seconds  
(without alignments)  
2758.288 Million cell updates/sec

Title: US-09-734-672-4  
Perfect score: 9649  
Sequence: 1 MDLSALRVEVQNVINAMQK.....LYCCQLDYVLPIQIPSHY 1863

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9649	100.0	1863	17 AAR81535	BRCA1 mutant from
2	9649	100.0	1863	19 AAR76099	Human BRCA1 (om12) p
3	9649	100.0	1863	21 AAR24219	Human BRCA1 (om13)
4	9642	99.9	1863	17 AAR97128	BRCA1, breast and
5	9642	99.9	1863	17 AAR81481	BRCA1, breast and
6	9642	99.9	1863	17 AAR91208	BRCA1, breast and
7	9642	99.9	1863	20 AAY32033	Human BRCA1 protei
8	9642	99.9	2353	22 ABO1696	Novel human diagno
9	9639	99.9	1863	17 AAR81490	BRCA1 mutant from
10	9639	99.9	1863	17 AAR81500	BRCA1 mutant from

11	9639	99.9	1863	17 AAR81522	BRCA1 mutant from
12	9639	99.9	1863	17 AAR81536	BRCA1 mutant from
13	9639	99.9	1863	17 AAR81540	BRCA1 mutant from
14	9638	99.9	1863	17 AAR81528	BRCA1 mutant from
15	9638	99.9	1863	17 AAR81532	BRCA1 mutant from
16	9638	99.9	1863	17 AAR81533	BRCA1 mutant from
17	9638	99.9	1863	17 AAR81534	BRCA1 mutant from
18	9638	99.9	1863	17 AAR81544	BRCA1 mutant from
19	9638	99.9	1863	17 AAR81545	BRCA1 mutant from
20	9638	99.9	1863	17 AAR81485	BRCA1 mutant from
21	9637	99.9	1863	17 AAR81493	BRCA1 mutant from
22	9637	99.9	1863	17 AAR81505	BRCA1 mutant from
23	9637	99.9	1863	17 AAR81511	BRCA1 mutant from
24	9637	99.9	1863	17 AAR81518	BRCA1 mutant from
25	9637	99.9	1863	17 AAR81538	BRCA1 mutant from
26	9636	99.9	1863	17 AAR81486	BRCA1 mutant from
27	9636	99.9	1863	17 AAR81509	BRCA1 mutant from
28	9636	99.9	1863	17 AAR81537	BRCA1 mutant from
29	9636	99.9	1863	17 AAR81542	BRCA1 mutant from
30	9636	99.9	1863	17 AAR81544	BRCA1 mutant from
31	9635	99.9	1863	17 AAR81524	BRCA1 mutant from
32	9635	99.9	1863	17 AAR81546	BRCA1 mutant from
33	9635	99.9	1863	18 AAW26522	Human BRCA1 consen
34	9635	99.9	1863	19 AAW79665	BRCA1 (om11) prote
35	9635	99.9	1863	19 AAW76100	Human BRCA1 om13 p
36	9635	99.9	1863	19 AAW76098	Human BRCA1 om11 p
37	9635	99.9	1863	21 AAB24217	Human BRCA1 (om11)
38	9635	99.9	1863	21 AAB24218	Human BRCA1 (om12)
39	9634	99.8	1863	17 AAR81488	BRCA1 mutant from
40	9634	99.8	1863	17 AAR81459	BRCA1 mutant from
41	9634	99.8	1863	17 AAR81539	BRCA1 mutant from
42	9630	99.8	1863	17 AAR89440	BRCA1 allele #8403
43	9630	99.8	1863	17 AAR81497	BRCA1 mutant from
44	9630	99.8	1863	18 AAM10003	Protein encoded by
45	9630	99.8	1863	19 AAW79889	Tumorigenic BRCA1

#### ALIGNMENTS

RESULT 1	
AAAR81535	AAAR81535 standard; Protein: 1863 AA.
ID	
XX	
AC	AAAR81535;
XX	
DT	02-OCT-1996 (first entry)
XX	
DE	BRCA1 mutant from PM06.
XX	
KW	Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW	antibody production; germline alteration; probe; lesion neoplasia; human;
KW	gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Misc-difference 871
FT	/note= "p871L"
XX	
PN	MO9605306-A2.
PD	22-FEB-1996.
XX	
PF	11-AUG-1995;
XX	
XX	95US-0483553.
PR	07-JUN-1995;
PR	12-AUG-1994;
PR	02-SEP-1994;
PR	16-SEP-1994;
PR	29-NOV-1994;
PR	24-MAR-1995;
PR	07-JUN-1995;
	95US-0480784.

XX (MYRI-) MYRIAD GENETICS INC.  
 PA (CANC-) CANCER INST.  
 PA (RECH-) CENT RECH DU CHUL.  
 XX  
 PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;  
 XX MPI: 1996-139702/14.  
 DR N-PSDB; AAT17493.  
 XX  
 PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers  
 XX  
 PS Claim 1: 218bp; English.  
 XX  
 CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein  
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)  
 CC (see AAR81481 for wild type protein). These mutations can be used as  
 CC immunogens for antibody production. The mutant BRCA1 genes encoding  
 CC these sequences have at least 1 mutation or polymorphism in comparison  
 CC to the wild type cDNA (see AAT17438 for wild type). By detecting a  
 CC germline alteration in the wild type BRCA1 gene, a predisposition for  
 CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA  
 CC isolated from a tissue sample from a subject has a probe, corresponding  
 CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an  
 CC allele-specific probe for a mutation of it), added to it. The conditions  
 CC allow for hybridization of the probe to the mRNA, and any hybridization  
 CC which occurs is detected. Alternatively the BRCA1 gene in the tissue  
 CC sample is isolated, and a shift in electrophoretic mobility of single  
 CC stranded DNA from the sample on a non-denaturing polyacrylamide gel  
 CC indicates a mutation. These methods of detection can also diagnose a  
 CC lesion neoplasia associated with the BRCA1 locus. The methods may be  
 CC used in gene therapy, protein replacement therapy and protein mimetics,  
 CC and may be used to screen for drugs in cancer therapy.  
 XX  
 XX Sequence 1863 AA:  
 SQ  
 Query Match 100.0%; Score 9649; DB 17; Length 1863;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY LIIIGAFVTEPOIIIOERPLTNKLRKRRTPTSGLPEDFIKKAADLAVOKTPMEMINOCTNOTE 540  
 DB LIIIGAFVTEPOIIIOERPLTNKLRKRRTPTSGLPEDFIKKAADLAVOKTPMEMINOCTNOTE 540  
 OY 541 QNGOVNMTNSGHNKTKGDSIONEKNPNDEISLEKESAFTKAEDISSISIMLELNI 600  
 DB 541 QNGOVNMTNSGHNKTKGDSIONEKNPNDEISLEKESAFTKAEDISSISIMLELNI 600  
 OY 601 HNSKAPKKNNLRKRSRTHIALELVYSRNLSPNCTEIQIDSCSSSEIKKKKNOMV 660  
 DB 601 HNSKAPKKNNLRKRSRTHIALELVYSRNLSPNCTEIQIDSCSSSEIKKKKNOMV 660  
 OY 661 RHSHNLOIMEGKEPATGAKKSNKPNEDQTSKRHSDPFPPELKLINAGSFTKSNTSELKE 720  
 DB 661 RHSHNLOIMEGKEPATGAKKSNKPNEDQTSKRHSDPFPPELKLINAGSFTKSNTSELKE 720  
 OY 721 FVNSLPREBEKEKLETVKVSNNNAEDPKDMLSGERYLOTERSVESSISLVPGTDTGYO 780  
 DB 721 FVNSLPREBEKEKLETVKVSNNNAEDPKDMLSGERYLOTERSVESSISLVPGTDTGYO 780  
 OY 781 ESISLLEVSTLGRAKTTPNNKCVSQCAFENPKCLIHGCSKDNNDTEGFKYPLGHEVNS 840  
 DB 781 ESISLLEVSTLGRAKTTPNNKCVSQCAFENPKCLIHGCSKDNNDTEGFKYPLGHEVNS 840  
 OY 841 RETSIEMESELDAOYLONTFKYKRSOFALFNPNGNAEBECATFSAHSGSLKKOSPKYT 900  
 DB 841 RETSIEMESELDAOYLONTFKYKRSOFALFNPNGNAEBECATFSAHSGSLKKOSPKYT 900  
 OY 901 FECEOKENEGKNESNIKRPQVTNITAGFPVVGOKRPVDMAKCSIKGSRCLLSQFNG 960  
 DB 901 FECEOKENEGKNESNIKRPQVTNITAGFPVVGOKRPVDMAKCSIKGSRCLLSQFNG 960  
 OY 961 NENGLITPNKHGLQNRYPRIPLPIKSPYKTKCKNLEENPEEHSNPEREMGENIP 1020  
 DB 961 NENGLITPNKHGLQNRYPRIPLPIKSPYKTKCKNLEENPEEHSNPEREMGENIP 1020  
 OY 1021 STVSTISRRNIRENVEKEASSNINEVSGSTNVEGSSINEIGSSDNIQAEGRNRPXL 1080  
 DB 1021 STVSTISRRNIRENVEKEASSNINEVSGSTNVEGSSINEIGSSDNIQAEGRNRPXL 1080  
 OY 1081 NAMRLGLVLOPEVYKOSLPQSNCKHPRIKKOEYEEVQVTNDFSPYLLISDMLQOPMGSS 1140  
 DB 1081 NAMRLGLVLOPEVYKOSLPQSNCKHPRIKKOEYEEVQVTNDFSPYLLISDMLQOPMGSS 1140  
 OY 1141 HASOVCSPTDLDLIDGELIKEDTSPFANDIKESSAVFSKSVQGELSRSPFTTHLQ 1200  
 DB 1141 HASOVCSPTDLDLIDGELIKEDTSPFANDIKESSAVFSKSVQGELSRSPFTTHLQ 1200  
 OY 1201 GYRGAKKLLESSEENLSEDEDELPCFOHLIFGVNNIPQOSTRHSTVATECLSKNTEENTL 1260  
 DB 1201 GYRGAKKLLESSEENLSEDEDELPCFOHLIFGVNNIPQOSTRHSTVATECLSKNTEENTL 1260  
 OY 1261 LSLKNSLINDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDITANTNTQDPFLIGS 1320  
 DB 1261 LSLKNSLINDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDITANTNTQDPFLIGS 1320  
 OY 1321 SKQMRHQSSEQVGLSDKELYSDDEBERGGLLENNQEBQSMNSNLGEAASGCESTSVSE 1380  
 DB 1321 SKQMRHQSSEQVGLSDKELYSDDEBERGGLLENNQEBQSMNSNLGEAASGCESTSVSE 1380  
 OY 1381 DCSGLSOSDILITPOORDTPQHNLIKLOOEAMLEVLAVLQHSOPSNSYPSITISSSALE 1440  
 DB 1381 DCSGLSOSDILITPOORDTPQHNLIKLOOEAMLEVLAVLQHSOPSNSYPSITISSSALE 1440  
 OY 1441 DLBNPEOSTSEKAVLITSQSSEYPISONPEGLISADKFEVYSAOSSTSKNKEPVEBSSPSK 1500  
 DB 1441 DLBNPEOSTSEKAVLITSQSSEYPISONPEGLISADKFEVYSAOSSTSKNKEPVEBSSPSK 1500  
 OY 1501 CPSLDRRWYMHSSGSLQNRNRPYQBELIKYVDVEEQOLEBEGPHDLTSTLTPRODLEG 1560  
 DB 1501 CPSLDRRWYMHSSGSLQNRNRPYQBELIKYVDVEEQOLEBEGPHDLTSTLTPRODLEG 1560

QY 1561 TPYLESGISLFSDDPESDPSERAPESARVGNIPSSSTALKVPDLKVAESASQSPAHAHTT 1620  
 DB 1561 TPYLESGISLFSDDPESDPSERAPESARVGNIPSSSTALKVPDLKVAESASQSPAHAHTT 1620  
 QY 1621 DTAGYANMESVSREKPELTASTERVNRKMSVYSGITPEEFMLVYKFAKHHITLTNLI 1680  
 DB 1621 DTAGYANMESVSREKPELTASTERVNRKMSVYSGITPEEFMLVYKFAKHHITLTNLI 1680  
 QY 1681 TEETHVVMKTDAEFVCEKTLKFLGAGKMWYVFWYQSIKERKMLNHEHPEVNGDY 1740  
 DB 1681 TEETHVVMKTDAEFVCEKTLKFLGAGKMWYVFWYQSIKERKMLNHEHPEVNGDY 1740  
 QY 1741 VNGRNHOGPKRARESDOKIFRGLGELICCYGPTNMTDLEMMVQLGASVARELSFTL 1800  
 DB 1741 VNGRNHOGPKRARESDOKIFRGLGELICCYGPTNMTDLEMMVQLGASVARELSFTL 1800  
 QY 1801 GTGVHPVYVQPAWMEEDNCFHAIIGMCEAPVYTRREVLDVALYQCELDITLIPQIP 1860  
 DB 1801 GTGVHPVYVQPAWMEEDNCFHAIIGMCEAPVYTRREVLDVALYQCELDITLIPQIP 1860  
 QY 1861 SHY 1863  
 DB 1861 SHY 1863

RESULT 2  
 AAM76099  
 ID AAM76099 standard; Protein; 1863 AA.

AC AAM76099;

DT 18-NOV-1998 (first entry)

DE Human BRCA1 cml2 protein.

KW BRCA1: cml2; human: breast and ovarian cancer predisposing gene;  
 KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 KW chromosome 17q.

OS Homo sapiens.

PN US5750400-A.

PD 12-MAY-1998.

PF 12-FEB-1997; 9705-0798691.

PR 12-FEB-1996; 9605-0598591.

PR 12-FEB-1997; 9705-0798691.

PA (ONCO-) ONCORMED INC.

PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;

PI Schelter DB, Zeng B;

DR WPI: 1998-296774/26.

DR N-PSDB; AAV46449.

PT BRCA1 cml2 gene coding sequences - useful for distinguishing between

PT polymorphisms and mutation(s) in the screening for disposition to

PT breast or ovarian cancer

XX disclosure; Column 47-56; 54pp; English.

XX This sequence represents the human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) cml2 gene. This sequence and polymorphic variations of  
 CC this sequence are useful for the identification of an individual who may  
 CC or may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome 17q  
 CC which is known to be linked to cancer susceptibility, especially breast

CC cancer. Cells containing a mutation in this gene lose the wild-type  
 CC function of BRCA1 and are more susceptible to cancers.

XX Sequence 1863 AA:

Query Match 100.0%; Score 9649; DB 19; Length 1863;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSALVEEYOVYINAMOKILBECTLELKEVSTKCHIRKFCFMLKLNOKKPSQ 60  
 DB 1 MDLSALVEEYOVYINAMOKILBECTLELKEVSTKCHIRKFCFMLKLNOKKPSQ 60  
 QY 61 CPLKNDITKRSLOESTRFQOLVBEELIKICAFOLDLGEVANSYNPAKKENSPBLKD 120  
 DB 61 CPLKNDITKRSLOESTRFQOLVBEELIKICAFOLDLGEVANSYNPAKKENSPBLKD 120  
 QY 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLVSQSLNGLCTVRTLRTKORIOPQTSVI 180  
 DB 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLVSQSLNGLCTVRTLRTKORIOPQTSVI 180  
 QY 181 ELGSDSEEDTVNKAATYCVGDDQLQTTPOGTREISLDSAKKACFSEDTVNTBHQ 240  
 DB 181 ELGSDSEEDTVNKAATYCVGDDQLQTTPOGTREISLDSAKKACFSEDTVNTBHQ 240  
 QY 241 PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVEPCGNTTHASSLOHENSLLTKDRMNV 300  
 DB 241 PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVEPCGNTTHASSLOHENSLLTKDRMNV 300  
 QY 301 KAEFCNKSQPGIARSOHNWAGSKETCNDRTPTSTKRYDLNADPLCEKREWNKOKLPC 360  
 DB 301 KAEFCNKSQPGIARSOHNWAGSKETCNDRTPTSTKRYDLNADPLCEKREWNKOKLPC 360  
 QY 361 SENPROTEDVPTLNTSSICKVNMFSRSDLLGSDSHGSEESNAKVAADVLDVNEVD 420  
 DB 361 SENPROTEDVPTLNTSSICKVNMFSRSDLLGSDSHGSEESNAKVAADVLDVNEVD 420  
 QY 421 EYSGSSSEKIDLASDPHEALICKSERVHSKSVESNIEDKIFGTYRKKAALPNLSHTEN 480  
 DB 421 EYSGSSSEKIDLASDPHEALICKSERVHSKSVESNIEDKIFGTYRKKAALPNLSHTEN 480  
 QY 481 LITGAFTPEOITIOERPLTNKLRKRRTSGLHPEDEIKRADLAVOKTPEMNOGTOTE 540  
 DB 481 LITGAFTPEOITIOERPLTNKLRKRRTSGLHPEDEIKRADLAVOKTPEMNOGTOTE 540  
 QY 541 QNGOVNMTNSGHENKTGDSIQNEKPNPIESLEKESAKTKTAEPITSSISNMELENT 600  
 DB 541 QNGOVNMTNSGHENKTGDSIQNEKPNPIESLEKESAKTKTAEPITSSISNMELENT 600  
 QY 601 HNSKAPKKNRLRRKSSSTRHIALELVYSRNLSPPNCTELOIDSCSSSEETKRRKYNOMP 660  
 DB 601 HNSKAPKKNRLRRKSSSTRHIALELVYSRNLSPPNCTELOIDSCSSSEETKRRKYNOMP 660  
 QY 661 RHSRNLQMEGKEPATGAKKSNKPNEDQSKRHSDTFPELKLTNAPGSFTKCSNTSELE 720  
 DB 661 RHSRNLQMEGKEPATGAKKSNKPNEDQSKRHSDTFPELKLTNAPGSFTKCSNTSELE 720  
 QY 721 FVNPSPREPKEKLELVKYSNNAEDPKDMLSGERLQTERVSESSISLVGTDTGTQ 780  
 DB 721 FVNPSPREPKEKLELVKYSNNAEDPKDMLSGERLQTERVSESSISLVGTDTGTQ 780  
 QY 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTGFYTPGLGHEVNH 840  
 DB 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTGFYTPGLGHEVNH 840  
 QY 841 RETSIEMESELDAQYLQNTFKVSKQSFALFSPNGNAEECATFSAHSGSLKQSPKVT 900  
 DB 841 RETSIEMESELDAQYLQNTFKVSKQSFALFSPNGNAEECATFSAHSGSLKQSPKVT 900  
 QY 901 FECEOKREENGKNESNITKPVQYVNTAGFPVYQOKKPVYNACSTIKGSRFLSSQFRG 960  
 DB 901 FECEOKREENGKNESNITKPVQYVNTAGFPVYQOKKPVYNACSTIKGSRFLSSQFRG 960

```

OY 961 NETGLITPNKHGILLQNPYRIPLLPFIKSFVKTKCKKNLLNEENDEHSHMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGILLQNPYRIPLLPFIKSFVKTKCKKNLLNEENDEHSHMSPEREMGNENIP 1020
OY 1021 STVSTISRNNINRENVFKKASSSNINEVSGSTNVEGSSINIEGSSDENIQAELOCRNRPKL 1080
DB 1021 STVSTISRNNINRENVFKKASSSNINEVSGSTNVEGSSINIEGSSDENIQAELOCRNRPKL 1080
OY 1081 NAMLRGVQPEVYKQSLPGSNCKHPEIKKOEYEEVQVNTPEFYLISDNEOPMGSS 1140
DB 1081 NAMLRGVQPEVYKQSLPGSNCKHPEIKKOEYEEVQVNTPEFYLISDNEOPMGSS 1140
OY 1141 HASQVCESTPDDLDDGEIKEDTSFAENDIKESSAAVFSKSVQKGLSRSPPTHTHLAQ 1200
DB 1141 HASQVCESTPDDLDDGEIKEDTSFAENDIKESSAAVFSKSVQKGLSRSPPTHTHLAQ 1200
OY 1201 GYRRAKKTLESSEENLSDEDELPQFHLFGKVNIPPSOSTHSTVAIECLSKNTEENL 1260
DB 1201 GYRRAKKTLESSEENLSDEDELPQFHLFGKVNIPPSOSTHSTVAIECLSKNTEENL 1260
OY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCELEDELTAANTQDPLIGS 1320
DB 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCELEDELTAANTQDPLIGS 1320
OY 1321 SKQMRHQSQGVGLSKDELVSDDERGTGLEENNOEBOGSDNSLGEAASGCESETSVSE 1380
DB 1321 SKQMRHQSQGVGLSKDELVSDDERGTGLEENNOEBOGSDNSLGEAASGCESETSVSE 1380
OY 1381 DCSGSSOSDILITTOORDTMQHNLIKLOQEMAELEVLROHGSQPSNSPTIISDSSALE 1440
DB 1381 DCSGSSOSDILITTOORDTMQHNLIKLOQEMAELEVLROHGSQPSNSPTIISDSSALE 1440
OY 1441 DLNRNEQSTSEKAVILTSOKSEYPISQNEBGLSADKFEVYSAADSTSKNKEPVERSSPSK 1500
DB 1441 DLNRNEQSTSEKAVILTSOKSEYPISQNEBGLSADKFEVYSAADSTSKNKEPVERSSPSK 1500
OY 1501 CPSLDDRWYMHSCSSGLQNRNRPQOEELIKYVDVEEQULEBSGPHDLTETSTYLRDLDLG 1560
DB 1501 CPSLDDRWYMHSCSSGLQNRNRPQOEELIKYVDVEEQULEBSGPHDLTETSTYLRDLDLG 1560
OY 1561 TPYLESGISLFSDDPESDPSEDRAPASAVGNIPSTSAKVPQLVAESAQSPAAAHHT 1620
DB 1561 TPYLESGISLFSDDPESDPSEDRAPASAVGNIPSTSAKVPQLVAESAQSPAAAHHT 1620
OY 1621 DTAGYNAMEESVSRKPELTASTERYNKRMSVVSGLTPEEFMLYKFKARKHHTITLNL 1680
DB 1621 DTAGYNAMEESVSRKPELTASTERYNKRMSVVSGLTPEEFMLYKFKARKHHTITLNL 1680
OY 1681 TEETHHYVMKTDAAEFVCERTLKYFLGIAGKMWVSIFWVTQSTKEKMLNEHDFEVRGV 1740
DB 1681 TEETHHYVMKTDAAEFVCERTLKYFLGIAGKMWVSIFWVTQSTKEKMLNEHDFEVRGV 1740
OY 1741 VNGRHHQGPKRARESDRKIFRGLGICCYGPTNMPDTDOLEWVQCGASVYKELSSFTL 1800
DB 1741 VNGRHHQGPKRARESDRKIFRGLGICCYGPTNMPDTDOLEWVQCGASVYKELSSFTL 1800
OY 1801 GTGVHPPIVVOGDANTEDNGFHAICOMCEAPVVTREWLDSVALYQOQELDTYVLIPOIFH 1860
DB 1801 GTGVHPPIVVOGDANTEDNGFHAICOMCEAPVVTREWLDSVALYQOQELDTYVLIPOIFH 1860
OY 1861 SHY 1863
DB 1861 SHY 1863

```

RESULT 3  
 AAB24219  
 ID AAB24219 standard: Protein: 1863 AA.  
 XX AAB24219:  
 XX 07-FEB-2001 (first entry)  
 XX

```

DE Human BRCA1 (om13) protein sequence SEQ ID NO:6.
XX
XX Human: BRCA1: chromosome 17: 17q21; breast cancer; ovarian cancer;
KW gene therapy; diagnosis; cytostatic; genetic susceptibility; mutation;
KW polymorphism; identification.
OS Homo sapiens.
XX
XX US6130322-A.
XX
XX 10-OCT-2000.
XX
XX 06-MAY-1998: 980S-0074476.
XX
XX 12-FEB-1996: 960S-0598591.
XX 12-DEC-1997: 970S-0798691.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Zeng B, Thurbur D, Olson SJ, Alvares CP, Allen ACP, Murphy PD:
PI Critz BS:
XX
XX WPI: 2000-646756/62.
XX
XX N-PSDB: AAC60795.
XX
XX Example 4: Column 65-74; 56pp; English.
XX
XX AAC60793 to AAC60795 encode the human BRCA1 (om1-3) proteins given in
CC AAB24217 to AAB24219 respectively. BRCA1 is found on chromosome 17
CC mapping to position 17q21. The BRCA1 (om12) coding sequence is
CC specifically claimed in the present invention. The BRCA1 (om12) coding
CC sequence is useful in gene therapy, especially for preventing or treating
CC breast or ovarian cancer. It is also useful for diagnosing or monitoring
CC breast or ovarian cancer. Furthermore, the BRCA1 (om12) coding sequence
CC is useful for: (a) identifying individuals having BRCA1 gene mutations
CC and having an increased genetic susceptibility to breast or ovarian
CC cancer, or identifying a mutation that increases the genetic
CC susceptibility to breast or ovarian cancer; (b) avoiding
CC misinterpretation of polymorphisms found in the BRCA1 gene; (c)
CC determining the presence of a previously unknown mutation in the BRCA1
CC gene; (d) probing a human sample of the BRCA1 gene by allele to determine
CC the presence of either polymorphic alleles or mutations; and (e)
CC performing diagnosis with a reagent derived from the BRCA1 (om1) cDNA
CC sequence. AAC60796 to AAC60861 represent PCR primers for the BRCA1 gene,
CC which are used in an example from the present invention.
XX
XX
SQ Sequence 1863 AA:
Query Match 100.0%; Score 9649; DB 21; Length 1863;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MDLSALRVEEYQNVINAMOKILECPICTLELKEPSTKCDHIFCFMLKLLNKKGPSQ 60
DB 1 MDLSALRVEEYQNVINAMOKILECPICTLELKEPSTKCDHIFCFMLKLLNKKGPSQ 60
OY 61 CPLCKNDITKRSLOESTRESQLVEELKTIICAFQDLDTGLEANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRESQLVEELKTIICAFQDLDTGLEANSYNFAKKENNSPEHLKD 120
OY 121 EVSIIQSMGYRNRRARKLQSEPNLSQETSLSVQLSNGLGYRTLRTRQRIQPKTSYI 180
DB 121 EVSIIQSMGYRNRRARKLQSEPNLSQETSLSVQLSNGLGYRTLRTRQRIQPKTSYI 180
OY 181 ELGSDSSEDTVKKATYCSVGOEQLQITPOGRDEISLDSAKKAACESEFDVNTTEHQ 240
DB 181 ELGSDSSEDTVKKATYCSVGOEQLQITPOGRDEISLDSAKKAACESEFDVNTTEHQ 240

```



QY 241 PSNDLNTTEKRAAEHRPEKYQSSVSNLHVEPCGTNTHASSLOHENSLLITKDRMNV 300  
 DB 241 PSNDLNTTEKRAAEHRPEKYQSSVSNLHVEPCGTNTHASSLOHENSLLITKDRMNV 300  
 QY 301 KAECNKSQOGLARSOHNRWAGSKETCNDRPTSTETKRYDLNDPLCEREMKOKLPC 360  
 DB 301 KAECNKSQOGLARSOHNRWAGSKETCNDRPTSTETKRYDLNDPLCEREMKOKLPC 360  
 QY 361 SENPRDEDPWMTLNSIOKVNMFPSPDELLSDSDSHOSESNAKVADVLNVEVD 420  
 DB 361 SENPRDEDPWMTLNSIOKVNMFPSPDELLSDSDSHOSESNAKVADVLNVEVD 420  
 QY 421 EYSSGSKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRRKASLPLNSHVTEN 480  
 DB 421 EYSSGSKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRRKASLPLNSHVTEN 480  
 QY 481 LIIAFTPEPOITIOERPLTNKLRKRRTSGLHPEDEITKADLAVOQTPEKINGTNOTE 540  
 DB 481 LIIAFTPEPOITIOERPLTNKLRKRRTSGLHPEDEITKADLAVOQTPEKINGTNOTE 540  
 QY 541 QNGOVANITNSGHEKNTKQDSIQNEKNPNPIESLEKESAFKTKAEPISSTISNNELENTI 600  
 DB 541 QNGOVANITNSGHEKNTKQDSIQNEKNPNPIESLEKESAFKTKAEPISSTISNNELENTI 600  
 QY 601 HNSKAPKKNRLRKSSTRHIALELVVSRNLSPNCTELQIDSCSSSEETIKKKKYNOMPV 660  
 DB 601 HNSKAPKKNRLRKSSTRHIALELVVSRNLSPNCTELQIDSCSSSEETIKKKKYNOMPV 660  
 QY 661 RHSNLIOLMGEKEPATATKAKSKNKNPNEOTSKRHSDTPELKLTAAPRSFTKCSMTSELKE 720  
 DB 661 RHSNLIOLMGEKEPATATKAKSKNKNPNEOTSKRHSDTPELKLTAAPRSFTKCSMTSELKE 720  
 QY 721 FVNPSPREKEEKELETVKVSNNNAEDPKDMLSGERVLQTERSVSSSISLVPSTDTGQTO 780  
 DB 721 FVNPSPREKEEKELETVKVSNNNAEDPKDMLSGERVLQTERSVSSSISLVPSTDTGQTO 780  
 QY 781 ESISLLEVSTLGAKTPEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLIGHEVNH 840  
 DB 781 ESISLLEVSTLGAKTPEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLIGHEVNH 840  
 QY 841 RETSIEMEESBLDOYLONTFKVSKROSFAFNSRNGAEECAATFSASHSGLKQSPKVT 900  
 DB 841 RETSIEMEESBLDOYLONTFKVSKROSFAFNSRNGAEECAATFSASHSGLKQSPKVT 900  
 QY 901 FECGOKRENGKNSNIKPQTVNITAGFPVYGOKDPVDNAKCSIKGSRFCLSSOPRG 960  
 DB 901 FECGOKRENGKNSNIKPQTVNITAGFPVYGOKDPVDNAKCSIKGSRFCLSSOPRG 960  
 QY 961 NETGLITPNKHGLQNDPYRIPLPEPIKSFVYTKCKKNLLENFEBHSMSPEREMGNENIP 1020  
 DB 961 NETGLITPNKHGLQNDPYRIPLPEPIKSFVYTKCKKNLLENFEBHSMSPEREMGNENIP 1020  
 QY 1021 STVSTIRNNIREVNEFEASSNINEVGSSTNEVGSSINELGSSDENIQAELGNNRGPKL 1080  
 DB 1021 STVSTIRNNIREVNEFEASSNINEVGSSTNEVGSSINELGSSDENIQAELGNNRGPKL 1080  
 QY 1081 NAMRLGLVLOPEYVKOSLPSNCKHPETIKOEVEVQVNTDPSPLISDNLQOPMGSS 1140  
 DB 1081 NAMRLGLVLOPEYVKOSLPSNCKHPETIKOEVEVQVNTDPSPLISDNLQOPMGSS 1140  
 QY 1141 HASOVCSPTDDLLDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSPPTHTHIAQ 1200  
 DB 1141 HASOVCSPTDDLLDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSPPTHTHIAQ 1200  
 QY 1201 GYRRGAKKLSSSEENLSSDEBELPCFQHLFGKYNNIPPSOSTRSTYATATECLSKNTEENL 1260  
 DB 1201 GYRRGAKKLSSSEENLSSDEBELPCFQHLFGKYNNIPPSOSTRSTYATATECLSKNTEENL 1260  
 QY 1261 LSLKNSLNDSCNOVYILAKASQEHHLSEETKCSASLFFSSQSELEDLTAANTTOOPFLIGS 1320  
 DB 1261 LSLKNSLNDSCNOVYILAKASQEHHLSEETKCSASLFFSSQSELEDLTAANTTOOPFLIGS 1320  
 QY 1321 SKOMRHOSQOGLSDKELVSDDERGTGLEENNOEQSMDSNLGEAASGCESETSVSE 1380

DB 1321 SKOMRHOSQOGLSDKELVSDDERGTGLEENNOEQSMDSNLGEAASGCESETSVSE 1380  
 QY 1381 DCSGLSSOSDILLTQOQDPMOHNLIKLOEAELEAVLEOHGSOPSNYPITISDSSALE 1440  
 DB 1381 DCSGLSSOSDILLTQOQDPMOHNLIKLOEAELEAVLEOHGSOPSNYPITISDSSALE 1440  
 QY 1441 DLRRPEOSTSEKAVLTQKSSSEYPISONPEGLISADKEFVSADSSSTSNKKEPVRRSSPSK 1500  
 DB 1441 DLRRPEOSTSEKAVLTQKSSSEYPISONPEGLISADKEFVSADSSSTSNKKEPVRRSSPSK 1500  
 QY 1501 CPSLDDRWYMHSCGSLQNRNYPQOEELIKVVDVEEQOLESGPHDLTETSYLRDQLEG 1560  
 DB 1501 CPSLDDRWYMHSCGSLQNRNYPQOEELIKVVDVEEQOLESGPHDLTETSYLRDQLEG 1560  
 QY 1561 TPYLESGISLFPDPEPSPEDRAPEARVGNITSSSTALKVPOLKVAESQSPAANTT 1620  
 DB 1561 TPYLESGISLFPDPEPSPEDRAPEARVGNITSSSTALKVPOLKVAESQSPAANTT 1620  
 QY 1621 DTAGYNAMESVSREKEBELTASTERVNKRMSMNVSGLTPEEFMLVYFARKHITLNL 1680  
 DB 1621 DTAGYNAMESVSREKEBELTASTERVNKRMSMNVSGLTPEEFMLVYFARKHITLNL 1680  
 QY 1681 TEETTHVVMKTDAEFVCEERTIKYFLGIAGCKWVSVYFWVTQSIKERKMLNEHDEFEVGDV 1740  
 DB 1681 TEETTHVVMKTDAEFVCEERTIKYFLGIAGCKWVSVYFWVTQSIKERKMLNEHDEFEVGDV 1740  
 QY 1741 VNGRNHOGPKRARRESORRTFRGLEICCYGFTWMPPTDQLEMMVOLGASVYKELSSFTL 1800  
 DB 1741 VNGRNHOGPKRARRESORRTFRGLEICCYGFTWMPPTDQLEMMVOLGASVYKELSSFTL 1800  
 QY 1801 GTGVHPIVVQPDAMTEWNGFHAIGOMCEAPVTVRENVLDVALYOCQELDTYLIPIPH 1860  
 DB 1801 GTGVHPIVVQPDAMTEWNGFHAIGOMCEAPVTVRENVLDVALYOCQELDTYLIPIPH 1860  
 QY 1861 SHY 1863  
 DB 1861 SHY 1863  
 RESULT 4  
 AAR97128  
 ID AAR97128 standard; Protein: 1863 AA.  
 XX AAR97128;  
 AC AAR97128;  
 DT 19-NOV-1996 (first entry)  
 XX  
 DE BRCA1, breast and ovarian cancer susceptibility gene product.  
 KW BRCA1; breast cancer; ovary cancer; predisposing gene; diagnosis; susceptibility gene; prognosis; gene therapy.  
 OS Homo sapiens.  
 PN M09605308-A1.  
 PD 22-FEB-1996.  
 XX  
 PF 11-AUG-1995; 95WO-0510220.  
 XX  
 PR 07-JUN-1995; 95US-0488011.  
 PR 12-AUG-1994; 94US-0289221.  
 PR 02-SEP-1994; 94US-0300266.  
 PR 16-SEP-1994; 94US-0308104.  
 PR 29-NOV-1994; 94US-0348824.  
 PR 24-MAR-1995; 95US-0409305.  
 PR 07-JUN-1995; 95US-0483554.  
 PR 07-JUN-1995; 95US-0487002.  
 XX  
 PA (MYRI-) MYRIAD GENETICS INC.  
 PA (UTAH) UNIV UTAH RES FOUND.  
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Futreal PA, Goldgar DE, Harshman KD, Kamb A, Miki Y;  
 PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;  
 PI Wiseman RM;  
 DR WPI, 1996-139704/14.  
 DR N-PSDB; AAT32601.  
 XX  
 PT New method for diagnosing a predisposition to breast and ovarian  
 PT cancer - by detecting a germline alteration in the BRCA1 gene or  
 PT gene regulatory sequence; for gene therapy and to screen for drugs  
 XX  
 PS Claim 12: Page 108-117; 200pp: English.

This is the amino acid sequence of the protein encoded by the breast and  
 CC ovarian cancer susceptibility gene, BRCA1. Four kindred families  
 CC provided genetic evidence for localisation of BRCA1 to a sufficiently  
 CC small region for the application of positional cloning strategies. A  
 CC detailed map of transcripts was developed for the region of 17q21 between  
 CC D17S1321 and D17S1324. A combination of sequences obtd. from cDNA  
 CC clones, hybrid-selected sequences and PCR prods. allowed construction of  
 CC a composite full-length BRCA1 cDNA (see AAT32612 for genomic sequence).  
 CC The BRCA1 cDNA revealed an open reading frame encoding a protein of  
 CC 208 KD. The protein also contains a sequence near the amino terminus  
 CC which had homology to zinc finger domains, esp. the sequence contains  
 CC Cys and His residues present as a C3HC4 zinc finger consensus motif.  
 CC The isolated cDNA is used in methods for either diagnosis of the  
 CC predisposition to cancer (partic. breast and ovarian cancer), or for the  
 CC diagnosis or prognosis of cancer, and also in gene-based therapies  
 CC directed at cancer cells.

XX Sequence 1863 AA:

Query Match 99.9%; Score 9642; DB 17; Length 1863;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEYOVNINAMOKIIECPICILELIKPEVSTKCDHIFCKECMLLNQKKGPSQ 60  
 DB 1 MDLSALRVEEYOVNINAMOKIIECPICILELIKPEVSTKCDHIFCKECMLLNQKKGPSQ 60  
 QY 61 CPLCKNDITKRSLQESTRESQVLEELKTIICAFQDGTGEYANSYNPAKKENNSPELKD 120  
 DB 61 CPLCKNDITKRSLQESTRESQVLEELKTIICAFQDGTGEYANSYNPAKKENNSPELKD 120  
 QY 121 EYSIIISMGYRNRAKRLQSEPNPSIQETSLSVQLSNLGTVRTLRKQIOPKTSVYI 180  
 DB 121 EYSIIISMGYRNRAKRLQSEPNPSIQETSLSVQLSNLGTVRTLRKQIOPKTSVYI 180  
 QY 181 ELGSDSSEDTVKNKATYCSVGDOELLQITPGTRDEISLDSAKKACGEFSETDVTNTEHQ 240  
 DB 181 ELGSDSSEDTVKNKATYCSVGDOELLQITPGTRDEISLDSAKKACGEFSETDVTNTEHQ 240  
 QY 241 PENNDLNTTEKRAEHRPEKYOGSSVSNLHVEPCGNTNTHASSIQHEHNSLLTKDRNVE 300  
 DB 241 PENNDLNTTEKRAEHRPEKYOGSSVSNLHVEPCGNTNTHASSIQHEHNSLLTKDRNVE 300  
 QY 301 KAEPCKSKOPGLARQOHNRWAGSKETCNDRTPTSTEEKKYVDLNAADPLCEKKNKOKLPC 360  
 DB 301 KAEPCKSKOPGLARQOHNRWAGSKETCNDRTPTSTEEKKYVDLNAADPLCEKKNKOKLPC 360  
 QY 361 SENPRDTEVPWITLNSIQKVNEMFSRDELLGSDSHDGESESNAKVADVLDVNEVD 420  
 DB 361 SENPRDTEVPWITLNSIQKVNEMFSRDELLGSDSHDGESESNAKVADVLDVNEVD 420  
 QY 421 EYSSGSSEKDLASDPHEALICKSERVHSHKSVSNIEDKIFGKTYRKASLPLMSHYTEN 480  
 DB 421 EYSSGSSEKDLASDPHEALICKSERVHSHKSVSNIEDKIFGKTYRKASLPLMSHYTEN 480  
 QY 481 LITGAFVTEPQITQIERPLTNKLKRRKRPSTGLHPEDFIKKAOLAVOKTPMIMQGNQTE 540  
 DB 481 LITGAFVTEPQITQIERPLTNKLKRRKRPSTGLHPEDFIKKAOLAVOKTPMIMQGNQTE 540

QY 541 QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSIMLELNI 600  
 DB 541 QNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSIMLELNI 600  
 QY 601 HNSKAPKKNNRLRRKSSRTHIALELVYSRNLSPNCTELQIDSCSSEIIRKKKYNOMPV 660  
 DB 601 HNSKAPKKNNRLRRKSSRTHIALELVYSRNLSPNCTELQIDSCSSEIIRKKKYNOMPV 660  
 QY 661 RHRNQLQMGKEPAGCAKSNKPNQOTSRHRSDTFPEELKLTJNAPGSTRKSNSELKE 720  
 DB 661 RHRNQLQMGKEPAGCAKSNKPNQOTSRHRSDTFPEELKLTJNAPGSTRKSNSELKE 720  
 QY 721 FVNPSPJPREKEEKLTVKVSNNAEDEPKDMLSGERVLOTERSVESSISLVPGDYGTQ 780  
 DB 721 FVNPSPJPREKEEKLTVKVSNNAEDEPKDMLSGERVLOTERSVESSISLVPGDYGTQ 780  
 QY 781 ESISLLEVSTLGAKAPENPKVCSQCAFENPKGLIHGSKDNRPDTGEFKYPLGHEVMS 840  
 DB 781 ESISLLEVSTLGAKAPENPKVCSQCAFENPKGLIHGSKDNRPDTGEFKYPLGHEVMS 840  
 QY 841 RETSIEMEESELDAQYLQNTFFKYSKRSFALSNPCNMAEECATSNAHSGSLKQSPRYT 900  
 DB 841 RETSIEMEESELDAQYLQNTFFKYSKRSFALSNPCNMAEECATSNAHSGSLKQSPRYT 900  
 QY 901 FECEOKEEENGNKNEINIKPQIVNITAGPVVQKDPVDNAKCSIIGSGRFLCSSQPRG 960  
 DB 901 FECEOKEEENGNKNEINIKPQIVNITAGPVVQKDPVDNAKCSIIGSGRFLCSSQPRG 960  
 QY 961 NETGLITPNKHGLQNPYRIPLEPIKSFVKTKCKKNLLENPEEHSHMSPEREMENIP 1020  
 DB 961 NETGLITPNKHGLQNPYRIPLEPIKSFVKTKCKKNLLENPEEHSHMSPEREMENIP 1020  
 QY 1021 STVSTISRRNIRENVKREASSNINEVSGSTNEVGSINIEIGSSPENIOALGRNRGKL 1080  
 DB 1021 STVSTISRRNIRENVKREASSNINEVSGSTNEVGSINIEIGSSPENIOALGRNRGKL 1080  
 QY 1081 NAMLRGLVLOPEVYKOSLPSCNCKHPEIKKQEEVEVYQNTDFSPYLLSDNLEQPMSS 1140  
 DB 1081 NAMLRGLVLOPEVYKOSLPSCNCKHPEIKKQEEVEVYQNTDFSPYLLSDNLEQPMSS 1140  
 QY 1141 HASQVSETPDDLLDQGEIKEDTSFAENDIKSSAFAFSKVGKGLSRSPFTTHLQ 1200  
 DB 1141 HASQVSETPDDLLDQGEIKEDTSFAENDIKSSAFAFSKVGKGLSRSPFTTHLQ 1200  
 QY 1201 GYRGAKKLLESSSEENLSEDEELPCQHLFGVNNIPESQSTRHSTVATECSKTEENL 1260  
 DB 1201 GYRGAKKLLESSSEENLSEDEELPCQHLFGVNNIPESQSTRHSTVATECSKTEENL 1260  
 QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEFTKCSASLFSQCSLELDLJANTNTQDPFLIGS 1320  
 DB 1261 LSLKNSLNDCSNOVILAKASQEHHLSEFTKCSASLFSQCSLELDLJANTNTQDPFLIGS 1320  
 QY 1321 SKQMRHQSOGVGLSDKELVSDDERGTCGLEENNOEDBSMNSNIGEAASGCESTSYSE 1380  
 DB 1321 SKQMRHQSOGVGLSDKELVSDDERGTCGLEENNOEDBSMNSNIGEAASGCESTSYSE 1380  
 QY 1381 DCSGLSSQSDILITTOORDIMOHNLKLOQEMAELEAVLEOHOSQPSNSYPIISQSSALE 1440  
 DB 1381 DCSGLSSQSDILITTOORDIMOHNLKLOQEMAELEAVLEOHOSQPSNSYPIISQSSALE 1440  
 QY 1441 DLRNPOSTSEKAVLTSOKSSEYPISQNPDEGLSADKFEVSADSSYTKKREVERSSPSK 1500  
 DB 1441 DLRNPOSTSEKAVLTSOKSSEYPISQNPDEGLSADKFEVSADSSYTKKREVERSSPSK 1500  
 QY 1501 CPSLDRWYMHSGSGSLQNRNTPSQQPELLKYVDVEEQULEESGPHDLTETSLPQDLEG 1560  
 DB 1501 CPSLDRWYMHSGSGSLQNRNTPSQQPELLKYVDVEEQULEESGPHDLTETSLPQDLEG 1560  
 QY 1561 TPYLESGISLFSDDPSPDEPRAPEASRVNIPSTSLAKVPQKVAESQSPAAAHYT 1620  
 DB 1561 TPYLESGISLFSDDPSPDEPRAPEASRVNIPSTSLAKVPQKVAESQSPAAAHYT 1620  
 QY 1621 DTAGYNAMEESVSRKPELJASTERVKNRSMVNSGLTPPEEFMLVYKPARRHHITLTNLI 1680

Db 1621 DTAGNAMESVSREKPELTAFTERVNRKMSMVVSGLTPEEFMLVYKFAKHHTLTNLI 1660  
 QY 1681 TEETHVVMKTDAFVCEERTLKTYFLGJAGKMWVSFYWMYOSIKERKMLNDEHVEGV 1740  
 Db 1681 TEETHVVMKTDAFVCEERTLKTYFLGJAGKMWVSFYWMYOSIKERKMLNDEHVEGV 1740  
 QY 1741 VNGRNHGGPKRARESDQRKIFRGLEICCYGPTNMPDQLEMMVQLGASVVKELSSFTL 1800  
 Db 1741 VNGRNHGGPKRARESDQRKIFRGLEICCYGPTNMPDQLEMMVQLGASVVKELSSFTL 1800  
 QY 1801 GTGCHPIVYVQPDAMTENDNGFHAIGOMCEAPVYTVRENVLDLSVALYQCELDITYLIPDIP 1860  
 Db 1801 GTGCHPIVYVQPDAMTENDNGFHAIGOMCEAPVYTVRENVLDLSVALYQCELDITYLIPDIP 1860  
 QY 1861 SHY 1863  
 Db 1861 SHY 1863  
 QY 1861 SHY 1863

RESULT 5  
 AAR81481  
 ID AAR81481 standard; Protein; 1863 AA.

AC AAR81481;  
 DT 30-SEP-1996 (first entry)  
 DE BRCA1.

KW Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;  
 KM antibody production; germline alteration; probe; lesion neoplasia; human;  
 KM gene therapy: protein replacement therapy; protein mimetic; BRCA1.

OS Homo sapiens.  
 PN W09605306-A2.  
 PD 22-FEB-1996.  
 XX 11-AUG-1995; 95MO-US10202.  
 PF 07-JUN-1995; 9505-0483553.  
 PR 12-AUG-1994; 9405-0289221.  
 PR 02-SEP-1994; 9405-0300266.  
 PR 16-SEP-1994; 9405-0308104.  
 PR 29-NOV-1994; 9405-0348824.  
 PR 24-MAR-1995; 9505-0409305.  
 PR 07-JUN-1995; 9505-0480784.

XX (MYRI-) MYRIAD GENETICS INC.  
 PA (CANC-) CANCER INST.  
 PA (RECH-) CENT RECH DU CHUL.

PI Durocher F, Eml M, Nakamura Y, Simard J, Shattuck-Eidens DM;  
 XX WPI; 1996-139702/14.  
 DR N-PSDB; AAT17438.  
 XX

PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers

PS Claim 1; Page 119-128; 218pp; English.

CC This sequence represents the protein encoded by the human breast and  
 CC ovarian cancer predisposing gene (BRCA1). Mutations of this sequence  
 CC (see AAR81483-R81497 and AAR81499-R81546) can be used as immunogens for  
 CC antibody production. The mutant BRCA1 genes have at least 1 mutation or  
 CC polymorphism in comparison to the cDNA encoding this sequence. By  
 CC detecting a germline alteration in this gene, a predisposition for  
 CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA  
 CC isolated from a tissue sample from a subject has a probe, corresponding

CC to a fragment of the cDNA encoding this sequence (or an allele-specific  
 CC probe for a mutation of it sequence), added to it. The conditions allow  
 CC for hybridisation of the probe to the mRNA, and any hybridisation which  
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample  
 CC is isolated, and a shift in electrophoretic mobility of single stranded  
 CC DNA from the sample on a non-denaturing polyacrylamide gel indicates a  
 CC mutation. These methods of detection can also diagnose a lesion  
 CC neoplasia associated with the BRCA1 locus. The methods may be used in  
 CC gene therapy, protein replacement therapy and protein mimetics, and may  
 CC be used to screen for drugs in cancer therapy.

XX Sequence 1863 AA:

Query: Match 99.9%; Score 9642; DB 17; Length 1863;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEYQVYINAMOKTLECPICLELKEPVSTKCDHIFCKFCMLKLNOKKPSQ 60  
 Db 1 MDLSALRVEEYQVYINAMOKTLECPICLELKEPVSTKCDHIFCKFCMLKLNOKKPSQ 60  
 QY 61 CPLCKNDITKRSLOESTRFSQVLELTKICAPOLDTGLEYANSYNFAKKENNSPEHLKD 120  
 Db 61 CPLCKNDITKRSLOESTRFSQVLELTKICAPOLDTGLEYANSYNFAKKENNSPEHLKD 120  
 QY 121 EVSIIQSMGYNRNRAKRLQSEPNPSLOETSLSYQLSNLGTVRTLRKRIQPKQTSYVI 180  
 Db 121 EVSIIQSMGYNRNRAKRLQSEPNPSLOETSLSYQLSNLGTVRTLRKRIQPKQTSYVI 180  
 QY 181 ELGSDSSEDTVANKATYCSVGQDELQITPQGTROEISLDSAKKACERSETVTNTEHHQ 240  
 Db 181 ELGSDSSEDTVANKATYCSVGQDELQITPQGTROEISLDSAKKACERSETVTNTEHHQ 240  
 QY 181 ELGSDSSEDTVANKATYCSVGQDELQITPQGTROEISLDSAKKACERSETVTNTEHHQ 240  
 Db 181 ELGSDSSEDTVANKATYCSVGQDELQITPQGTROEISLDSAKKACERSETVTNTEHHQ 240  
 QY 241 PSNNDLMTTEKRAERHPEKYQGSVSNLHVEPCGTTHASSLOHENSLLTTRDNMVE 300  
 Db 241 PSNNDLMTTEKRAERHPEKYQGSVSNLHVEPCGTTHASSLOHENSLLTTRDNMVE 300  
 QY 301 KAFCNKSQKQGLARSOHNWAGSKETCNDRTPTSTCKKYDLNADPLCERKENNOKLPC 360  
 Db 301 KAFCNKSQKQGLARSOHNWAGSKETCNDRTPTSTCKKYDLNADPLCERKENNOKLPC 360  
 QY 361 SENPDTEDEVWITLANSIOKVNEMFSRSDLLGSDSDSHGSESNKAVADVLDVLEND 420  
 Db 361 SENPDTEDEVWITLANSIOKVNEMFSRSDLLGSDSDSHGSESNKAVADVLDVLEND 420  
 QY 421 EYSSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRRKASLPLNISHVEN 480  
 Db 421 EYSSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRRKASLPLNISHVEN 480  
 QY 481 LIIGAFVTEPOIIOERPLTNKLRKRRPTSGIHPDEFIRKADLAVOKTPEKINOCTOTE 540  
 Db 481 LIIGAFVTEPOIIOERPLTNKLRKRRPTSGIHPDEFIRKADLAVOKTPEKINOCTOTE 540  
 QY 541 QNGOVNITNSGHNKTKGDSIONEKNPNPIESLEKESAKTYAEPITSSISNMELENI 600  
 Db 541 QNGOVNITNSGHNKTKGDSIONEKNPNPIESLEKESAKTYAEPITSSISNMELENI 600  
 QY 601 HNSKAPKKNRLRRSSSRHIALELYVSRNLSPNCNCELOIDSSSSEELIKKKYNMPV 660  
 Db 601 HNSKAPKKNRLRRSSSRHIALELYVSRNLSPNCNCELOIDSSSSEELIKKKYNMPV 660  
 QY 661 RHSRNLQIMEGKEPATGAKSKNPNEDOTSKRHSDTPPELKLINAPOSFTKCSNTSLKE 720  
 Db 661 RHSRNLQIMEGKEPATGAKSKNPNEDOTSKRHSDTPPELKLINAPOSFTKCSNTSLKE 720  
 QY 721 FVNPSPLEREKEKLETVKVSNNAEDEPKDMLSGERVLOTERSVSSSISLVBGTDVGTQ 780  
 Db 721 FVNPSPLEREKEKLETVKVSNNAEDEPKDMLSGERVLOTERSVSSSISLVBGTDVGTQ 780  
 QY 781 EESISLEVSTLGAKTEPKNKVSQCAAFENPKGLIHGCSKDNNDDEGFRYPLGHEVNH 840  
 Db 781 EESISLEVSTLGAKTEPKNKVSQCAAFENPKGLIHGCSKDNNDDEGFRYPLGHEVNH 840

```

QY 841 RETSIEMESELDAOYLQNTFKVSKROSFALEFSPNGAEECATFSAHSGSLKKQSPKV 900
DB 841 RETSIEMESELDAOYLQNTFKVSKROSFALEFSPNGAEECATFSAHSGSLKKQSPKV 900
QY 901 FECEQKEENQKESNINIKPVQVNTITAGFVVGOKDPVNNAKCSIGGSRFCLSSQFRG 960
DB 901 FECEQKEENQKESNINIKPVQVNTITAGFVVGOKDPVNNAKCSIGGSRFCLSSQFRG 960
QY 961 NETGLTPNKHGLQNPYRIPLEPIKSFYKTKCKKLLDENFEHSMSPEREKGNINIP 1020
DB 961 NETGLTPNKHGLQNPYRIPLEPIKSFYKTKCKKLLDENFEHSMSPEREKGNINIP 1020
QY 961 NETGLTPNKHGLQNPYRIPLEPIKSFYKTKCKKLLDENFEHSMSPEREKGNINIP 1020
DB 961 NETGLTPNKHGLQNPYRIPLEPIKSFYKTKCKKLLDENFEHSMSPEREKGNINIP 1020
QY 1021 STVSTISRNINIRENVEKEASSNINEVSGSTNEVGSSINEIGSSDENIQAELGNRPKL 1080
DB 1021 STVSTISRNINIRENVEKEASSNINEVSGSTNEVGSSINEIGSSDENIQAELGNRPKL 1080
QY 1081 NAMLRIGVLOPEYKOSLPKSNCKHPEIKOEYEVQVNTDPSPLIDNLEQPMGSS 1140
DB 1081 NAMLRIGVLOPEYKOSLPKSNCKHPEIKOEYEVQVNTDPSPLIDNLEQPMGSS 1140
QY 1081 NAMLRIGVLOPEYKOSLPKSNCKHPEIKOEYEVQVNTDPSPLIDNLEQPMGSS 1140
DB 1081 NAMLRIGVLOPEYKOSLPKSNCKHPEIKOEYEVQVNTDPSPLIDNLEQPMGSS 1140
QY 1141 HASQVSETPDDLDDGEIKEDTSFAENDIKESSAVFSKVGKELSRSPFTHTHIAQ 1200
DB 1141 HASQVSETPDDLDDGEIKEDTSFAENDIKESSAVFSKVGKELSRSPFTHTHIAQ 1200
QY 1201 GYRGAKKLESSEENLSEDELLPCFQHLFGVNNIPSGSTRHSTVATECLSKNTEBL 1260
DB 1201 GYRGAKKLESSEENLSEDELLPCFQHLFGVNNIPSGSTRHSTVATECLSKNTEBL 1260
QY 1261 LSKNSLNOCNSNYIILAKASQEHHLSEETCSASLPSQSCSELEDLTANTNTQDPLIGS 1320
DB 1261 LSKNSLNOCNSNYIILAKASQEHHLSEETCSASLPSQSCSELEDLTANTNTQDPLIGS 1320
QY 1321 SKQMRHQSQOGVGLSDKELVSDDEERGTLLEENQEOGMDSNLGAASCESETVSE 1380
DB 1321 SKQMRHQSQOGVGLSDKELVSDDEERGTLLEENQEOGMDSNLGAASCESETVSE 1380
QY 1381 DCSGSSQSDILTTQORDTMOHNLKLOQMALEAVLEQHGQSPNSYPSIISDSALE 1440
DB 1381 DCSGSSQSDILTTQORDTMOHNLKLOQMALEAVLEQHGQSPNSYPSIISDSALE 1440
QY 1441 DLRNPQSTSEKAVLTSQKSEYPIQONPGLSADRFVADSDSTGNKEPVERSSPSK 1500
DB 1441 DLRNPQSTSEKAVLTSQKSEYPIQONPGLSADRFVADSDSTGNKEPVERSSPSK 1500
QY 1501 CFSLDDRWYMHSCSGSLQNRNYPQOELLKVVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560
DB 1501 CFSLDDRWYMHSCSGSLQNRNYPQOELLKVVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560
QY 1561 TPYLESIGLFSDDPSDESDRAPESARVGNIPSSSTSALKVPQLKVAESAQSPAAAHNT 1620
DB 1561 TPYLESIGLFSDDPSDESDRAPESARVGNIPSSSTSALKVPQLKVAESAQSPAAAHNT 1620
QY 1621 DRAGYVAMEESVSRKPELTASTERYNKRMSVVGSLTPPEFMVLVYFAKHHHTLNL 1680
DB 1621 DRAGYVAMEESVSRKPELTASTERYNKRMSVVGSLTPPEFMVLVYFAKHHHTLNL 1680
QY 1681 TEETHHVAKTAEFCERTLKYFLGJAGKMWVSYFWVYQSIKERKMLNEHDFEYAGDV 1740
DB 1681 TEETHHVAKTAEFCERTLKYFLGJAGKMWVSYFWVYQSIKERKMLNEHDFEYAGDV 1740
QY 1741 VNGRNHQGKRRARESODRKIFRGLEICCGPFTNMPDLEMMVQOLCGASVYKELSSFTL 1800
DB 1741 VNGRNHQGKRRARESODRKIFRGLEICCGPFTNMPDLEMMVQOLCGASVYKELSSFTL 1800
QY 1801 GTGVHPHYVVOADWTEGNGFAHIGOMCEAPVYTRRENVLDVALYQCOELDTYLIQPIH 1860
DB 1801 GTGVHPHYVVOADWTEGNGFAHIGOMCEAPVYTRRENVLDVALYQCOELDTYLIQPIH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863

```

```

RESULT 6
ID AAR91208
XX AAR91208 standard; Protein; 1863 AA.
AC AAR91208;
XX
DT 05-JUN-1996 (first entry)
XX
DE BRCA1..breast and ovarian cancer predisposing gene product.
XX
KW BRCA1..breast cancer; ovarian cancer; predisposing gene;
KW susceptibility gene; protein replacement therapy; diagnosis;
KW prognosis.
XX
OS Homo sapiens.
XX
PN W09605307-A2.
PD 22-FEB-1996.
XX
PE 11-AUG-1995; 95WO-US10203.
PR 07-JUN-1995; 95US-0488011.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0483554.
PR 07-JUN-1995; 95US-0487002.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (UTAH ) UNIT UTAH RES. FOUND.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Eutreal AP, Goldgar DE, Harsiman KD, Kamb A, Miki Y;
PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;
PI Wiseman RW;
XX
DR WPI: 1996-139703/14.
DR N-PSDB: AAT18310.
XX
PT New isolated human cancer predisposing gene, BRCA1 - used to develop
PT prods. for diagnosis, prognosis and therapy of cancers, partic.
PT breast and ovarian cancers
XX
PS Claim 1; Page 117-124; 190pp; English.
XX
CC The BRCA1 polypeptide (AAR91208) is the product of the human BRCA1
CC gene (AAT18310). It can be obtd. by insertion of BRCA1 DNA into a
CC vector and expression in host cells; products of mutated BRCA1
CC genes, associated in humans with a predisposition to breast and
CC ovarian cancer, can also be produced. BRCA1 polypeptides are used
CC as immunogens for the prodn. of antibodies, or to screen therapeutic
CC drugs. They can also be used in protein replacement therapy.
XX
SQ Sequence 1863 AA.
Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEVQVIVANOKILECPICLEIKPEVSTKCDHIECFCKMLKLNOKKGPSQ 60
DB 1 MDLSALRVEVQVIVANOKILECPICLEIKPEVSTKCDHIECFCKMLKLNOKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRESQVLEELKTTICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRESQVLEELKTTICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIOSMGYRNRAKRLLOSPEPNLSLOETSLSVQLSNUGYRTLRTRORIOPOKTSYVI 180
DB 121 EVSIIOSMGYRNRAKRLLOSPEPNLSLOETSLSVQLSNUGYRTLRTRORIOPOKTSYVI 180

```

QY 181 ELGSDSEDTVNKATYCSVODDELLQITPOGTREDEISLDSAKKACEFSETDVTNTEHHO 240  
 DB 181 ELGSDSEDTVNKATYCSVODDELLQITPOGTREDEISLDSAKKACEFSETDVTNTEHHO 240  
 QY 241 PSNNDLNTEKRAAEHHPKYOSSVNLHVEPCGTNTHASSLOHENSLLTKDRMAYE 300  
 DB 241 PSNNDLNTEKRAAEHHPKYOSSVNLHVEPCGTNTHASSLOHENSLLTKDRMAYE 300  
 QY 301 KAFCKSKQOPGLARQOHNRWAGSKETCDNRTPSTEEKYVDLADLCEKRNKKKLTFC 360  
 DB 301 KAFCKSKQOPGLARQOHNRWAGSKETCDNRTPSTEEKYVDLADLCEKRNKKKLTFC 360  
 QY 361 SENPRTDEVPMTLLNSSIOKVEMFSRDELLGSDSHDGESESNAKADVLDVNEVD 420  
 DB 361 SENPRTDEVPMTLLNSSIOKVEMFSRDELLGSDSHDGESESNAKADVLDVNEVD 420  
 QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKASLPNLSHVTEN 480  
 DB 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKASLPNLSHVTEN 480  
 QY 481 LIIGAVTEPQIIQIEPRLNKLKRRRPTSGLHPEDFIKADLAVOKTEPMINOGNOTE 540  
 DB 481 LIIGAVTEPQIIQIEPRLNKLKRRRPTSGLHPEDFIKADLAVOKTEPMINOGNOTE 540  
 QY 541 QNGOVNITNSGHEKTKGDSIONEKNPPIESLEKESAFKTKAEPJISSISIMLELNI 600  
 DB 541 QNGOVNITNSGHEKTKGDSIONEKNPPIESLEKESAFKTKAEPJISSISIMLELNI 600  
 QY 601 HNSKAPKRNLRKRSSTRHIALELVVSNLSPNCTELOIDSCSSSEIKKKKYNOMV 660  
 DB 601 HNSKAPKRNLRKRSSTRHIALELVVSNLSPNCTELOIDSCSSSEIKKKKYNOMV 660  
 QY 661 RUSRNQIMEGKEPAPGAKSKNPNQOTSKRHSODTFPELKLITNAGSTKCSNTSELKE 720  
 DB 661 RUSRNQIMEGKEPAPGAKSKNPNQOTSKRHSODTFPELKLITNAGSTKCSNTSELKE 720  
 QY 721 FVNPSLPREKEKEKLETVVSNNAEDPKDMLSGEVLQTERSVESSTISLVGTDTYGTQ 780  
 DB 721 FVNPSLPREKEKEKLETVVSNNAEDPKDMLSGEVLQTERSVESSTISLVGTDTYGTQ 780  
 QY 781 ESISLLEVSTLQKAKTEPNKCVSOCAAFENPKGLHGCSKDNBNDETEGFKYPLGHEVNS 840  
 DB 781 ESISLLEVSTLQKAKTEPNKCVSOCAAFENPKGLHGCSKDNBNDETEGFKYPLGHEVNS 840  
 QY 841 RETSIEMESELDAOYLONTFVSKROSALFSPNPNABEBCATPSAHSGLKKOSPKYT 900  
 DB 841 RETSIEMESELDAOYLONTFVSKROSALFSPNPNABEBCATPSAHSGLKKOSPKYT 900  
 QY 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDMNAKCSIKGSSRFLSSQFRG 960  
 DB 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDMNAKCSIKGSSRFLSSQFRG 960  
 QY 961 NETGLITPKHGLONPYRIPPLPIKSVKTKCKKNLLEENFEHSMSPEREMGENITP 1020  
 DB 961 NETGLITPKHGLONPYRIPPLPIKSVKTKCKKNLLEENFEHSMSPEREMGENITP 1020  
 QY 1021 STVSTISRNINIRENFKEASSNININVGSTINIEGSSIDENIOAELGRNGPXL 1080  
 DB 1021 STVSTISRNINIRENFKEASSNININVGSTINIEGSSIDENIOAELGRNGPXL 1080  
 QY 1081 NAMRLGVQPEVYKQSLPGSNCKHPEIKOEVEEVQTVNIDFSYLLSDMLEQMGSS 1140  
 DB 1081 NAMRLGVQPEVYKQSLPGSNCKHPEIKOEVEEVQTVNIDFSYLLSDMLEQMGSS 1140  
 QY 1141 HASOVCESETPDDLDDGETIKEDTSPAENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200  
 DB 1141 HASOVCESETPDDLDDGETIKEDTSPAENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200  
 QY 1201 GYRGAKKLESSEENLSEDELPCHOLLPFGKVVNNIPSOSTRHSVATECISKNTENTL 1260  
 DB 1201 GYRGAKKLESSEENLSEDELPCHOLLPFGKVVNNIPSOSTRHSVATECISKNTENTL 1260

QY 1261 LSLKNSLNDSCSNVILAKASOEHLSEETKCSASLFFSSQCELEDTLTANTNTODPELIGS 1320  
 DB 1261 LSLKNSLNDSCSNVILAKASOEHLSEETKCSASLFFSSQCELEDTLTANTNTODPELIGS 1320  
 QY 1321 SKOMRHQSESQGVGLSDKELVSDDEBRTGLEENNOEOMSNDNLGEAASCESETSYSE 1380  
 DB 1321 SKOMRHQSESQGVGLSDKELVSDDEBRTGLEENNOEOMSNDNLGEAASCESETSYSE 1380  
 QY 1381 DCSGLSSQSDILITTOOROTMOMHLIKLOQEMALEVLQDHGQSPNSYPSIISDSSALE 1440  
 DB 1381 DCSGLSSQSDILITTOOROTMOMHLIKLOQEMALEVLQDHGQSPNSYPSIISDSSALE 1440  
 QY 1441 DLNRPQOSTSEKAVLTSOKSSEYPIQONPEGLSADFEVSADSTSKNKEPVERSSPSK 1500  
 DB 1441 DLNRPQOSTSEKAVLTSOKSSEYPIQONPEGLSADFEVSADSTSKNKEPVERSSPSK 1500  
 QY 1501 CFSLDRLWYMHSCSGSLQONRNPPOBELIKVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560  
 DB 1501 CFSLDRLWYMHSCSGSLQONRNPPOBELIKVVDVEEQOLEESGPHDLTETSYLPRODLEG 1560  
 QY 1561 TPYLESGISLFDSDPESDEDRAPESARVNTIPSSTSALKVQLKVAESAQSPAANT 1620  
 DB 1561 TPYLESGISLFDSDPESDEDRAPESARVNTIPSSTSALKVQLKVAESAQSPAANT 1620  
 QY 1621 DTAGYNAMEESVSREKPELTASTERYNKRMSWVSGLTPEEPMLYKFKARKHITLTNLI 1680  
 DB 1621 DTAGYNAMEESVSREKPELTASTERYNKRMSWVSGLTPEEPMLYKFKARKHITLTNLI 1680  
 QY 1681 TEETHVVKKTDAEFVCERTLKYFLIGIAGKVVVSYFWVYOSIKERKMLNEHDFEVRGCV 1740  
 DB 1681 TEETHVVKKTDAEFVCERTLKYFLIGIAGKVVVSYFWVYOSIKERKMLNEHDFEVRGCV 1740  
 QY 1741 VNCRNHQGKRARESDRKIFRGLETICYGPTNMPTDOLLEWVQLCGASVKESSLFTL 1800  
 DB 1741 VNCRNHQGKRARESDRKIFRGLETICYGPTNMPTDOLLEWVQLCGASVKESSLFTL 1800  
 QY 1801 GTGVHPYVVPDAMWEDNGFHAIGOMCEAPVYTRFVLDVSVALYOCQELDTVYLIPQIPH 1860  
 DB 1801 GTGVHPYVVPDAMWEDNGFHAIGOMCEAPVYTRFVLDVSVALYOCQELDTVYLIPQIPH 1860  
 QY 1861 SHY 1863  
 DB 1861 SHY 1863  
 RESULT 7  
 AAY32033  
 ID AAY32033 standard; protein; 1863 AA.  
 AC AAY32033;  
 DT 05-JAN-2000 (first entry)  
 DE Human BRCA1 protein.  
 KW BRCA1; p53 protein; p21 gene; human; tumour suppressor;  
 KW transcriptional activator; breast cancer; cell proliferation;  
 KW apoptosis; diagnosis; anticancer; antitumour; drug screening.  
 OS Homo sapiens.  
 OS  
 FH Key location/Qualifiers  
 FT Binding-site 224..500  
 FT "note="binding site of p53 protein"  
 FT Region 499..510  
 FT "note="NLS region"  
 PN MO9950280-A1.  
 PD 07-OCT-1999.  
 PF 31-MAR-1999; 99MO-0507150.

PR	31-MAR-1998;	98US-0080146.
XX	(UYPE-) UNIV. PENNSYLVANIA.	
PA	El-Delfy WS, Weber BL;	
PI	WPI, 1999-601319/51.	
XX		
DR		
XX		
PT	Nucleic acid involved in BRCA-1-mediated control of transcriptional	
PT	regulation of tumour suppressor genes and related peptides, used to	
PT	screen for modulators for use as anticancer agents -	
PS	Claim 9; Page 76-80; 94pp; English.	
XX		
CC	This sequence represents the human BRCA1 protein. BRCA-1	
CC	inhibits S-phase cell-cycle progression by transactivating	
CC	expression of p21 in a p53-independent manner. It also	
CC	interacts with p53 (see AA12034) in vitro and in vivo, and	
CC	increases p53-dependent transcription from the p21 (see AA20131)	
CC	and bax promoters. These newly discovered pathways of BRCA1 action	
CC	affecting novel targets to which pharmaceutical agents capable of	
CC	affecting cell proliferation and apoptosis can be isolated, and	
CC	can then be used in the treatment and control of cellular	
CC	proliferation disorders, such as breast cancer. Assays and	
CC	compositions for identifying compounds that enhance or repress	
CC	cellular proliferation via these BRCA1-mediated pathways are	
CC	disclosed.	
CC		
xx		
xx		
SQ	Sequence 1863 AA;	
	Query Match 99.9%; Score 9642; DB 20; Length 1863;	
	Best Local Similarity 99.9%; Pred. No. 0;	
	Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0.	
OY	1 MDLSALRVEEYONVATNAOKITIECPICIELIKEPYSTCDHIFCFECFCHLKNQKRGSG	60
Dd	1 MDLSALRVEEYONVATNAOKITIECPICIELIKEPYSTCDHIFCFECFCHLKNQKRGSG	60
OY	61 CPLCNDLTTRKSLOESTFRFSOLVEELKTIICAFOLDTGLEYANSYNFAKKENSPHLKD	120
Dd	61 CPLCNDLTTRKSLOESTFRFSOLVEELKTIICAFOLDTGLEYANSYNFAKKENSPHLKD	120
OY	121 EVSIIOSMGYRNRAKRLLOSPEBNSLOETSLSVGLNIGVTPLRTQRIOPOKTSYVI	180
Dd	121 EVSIIOSMGYRNRAKRLLOSPEBNSLOETSLSVGLNIGVTPLRTQRIOPOKTSYVI	180
OY	181 ELGSSSDSDTYNKAKYCVSGOELLQITPPCGRIDISLDSAKKACSESDVTNTBHQ	240
Dd	181 ELGSSSDSDTYNKAKYCVSGOELLQITPPCGRIDISLDSAKKACSESDVTNTBHQ	240
OY	241 PSNNDLNTEKRAAERHPKEKYOGSSVSMLHYEPCCTNHASSIQHENSLLTTRDMAYE	300
Dd	241 PSNNDLNTEKRAAERHPKEKYOGSSVSMLHYEPCCTNHASSIQHENSLLTTRDMAYE	300
OY	301 KAFCCKNSKOPGLARSQHNRWAGSKETCNDRTPSTEEKVLDINADPLCEKREWNQKLPC	360
Dd	301 KAFCCKNSKOPGLARSQHNRWAGSKETCNDRTPSTEEKVLDINADPLCEKREWNQKLPC	360
OY	361 SENRPDPEADVMTILNSSIQKVNEFMSDELGSDDSHDESESNAAVAVDVLNEDV	420
Dd	361 SENRPDPEADVMTILNSSIQKVNEFMSDELGSDDSHDESESNAAVAVDVLNEDV	420
OY	421 EYSGSSEKIDLASDPHEALICKSERVASKSVESNIIEBKIGKTYRRKASLPNLISHVEN	480
Dd	421 EYSGSSEKIDLASDPHEALICKSERVASKSVESNIIEBKIGKTYRRKASLPNLISHVEN	480
OY	481 LIIGAFTYEPOIIIOERPLTNKLRKRREPISGLIHEDFIKADLAVALQKPEMINOQTOTE	540
Dd	481 LIIGAFTYEPOIIIOERPLTNKLRKRREPISGLIHEDFIKADLAVALQKPEMINOQTOTE	540
OY	541 QNGCYVMNITNSGHENKTKGDSIQNEKNPNPLESELEKSAFTKAAPLISSSTSNNLELNT	600
Dd	541 QNGCYVMNITNSGHENKTKGDSIQNEKNPNPLESELEKSAFTKAAPLISSSTSNNLELNT	600

QY	601	HNSKAKKKNRLRKKSSTRHIALEYVSRNLSPNCTELQIDSCSSSEETKKKKYQMAY	660
Db	601	HNSKAKKKNRLRKKSSTRHIALEYVSRNLSPNCTELQIDSCSSSEETKKKKYQMAY	660
QY	661	RHSRNLQWMBGKEPPATGAKKSNKPNMOTSRRHSDPFPELKLTJNAGSFFKCSNTSELKE	720
Db	661	RHSRNLQWMBGKEPPATGAKKSNKPNMOTSRRHSDPFPELKLTJNAGSFFKCSNTSELKE	720
QY	721	FVNPSLPREKEBEKLETVAVSNNAEDPKDLMJGSEVYLQTERSVESSISLVPCTDYGTQ	780
Db	721	FVNPSPREKEBEKLETVAVSNNAEDPKDLMJGSEVYLQTERSVESSISLVPCTDYGTQ	780
QY	781	ESTSLLEVTSTCKAKTEPNKCVSQAAPFNPGCLTHGCSKDNNDPTEGKRYPLGHEVHNS	840
Db	781	ESTSLLEVTSTCKAKTEPNKCVSQAAPFNPGCLTHGCSKDNNDPTEGKRYPLGHEVHNS	840
QY	841	RETSLEMESELDQAQYLOMTFVYSKRSOFALSNPONAEECATFSAHSGSLKKOSPKYT	900
Db	841	RETSLEMESELDQAQYLOMTFVYSKRSOFALSNPONAEECATFSAHSGSLKKOSPKYT	900
QY	901	FECBOKEENQGNESNIKPVQYNTTAGPPVYGQDKPYDNAKCSITKGSRCFLSSQFRG	960
Db	901	FECBOKEENQGNESNIKPVQYNTTAGPPVYGQDKPYDNAKCSITKGSRCFLSSQFRG	960
QY	961	NETGLTTPMKHGLLOPNRYIPLPLFTJPSVTKCKKNLLEENPEEHSMPREDMCENETP	1020
Db	961	NETGLTTPMKHGLLOPNRYIPLPLFTJPSVTKCKKNLLEENPEEHSMPREDMCENETP	1020
QY	1021	STVSTSRNNIRENVEKFASSNNINEVGSSTNEVGSSINETIGSSDENIOAELGRNKPXL	1080
Db	1021	STVSTSRNNIRENVEKFASSNNINEVGSSTNEVGSSINETIGSSDENIOAELGRNKPXL	1080
QY	1081	NAMLRLGVLOPEYKOSILPGNSCKHPEIKKOYEEOVQYVNTDFSPYLLSDMLEOPMGSS	1140
Db	1081	NAMLRLGVLOPEYKOSILPGNSCKHPEIKKOYEEOVQYVNTDFSPYLLSDMLEOPMGSS	1140
QY	1141	HASQVCSFPPDLDLDGELIKEDTSAENDIKESSAVFYSKVQGBELSRSPFPTHHLAQ	1200
Db	1141	HASQVCSFPPDLDLDGELIKEDTSAENDIKESSAVFYSKVQGBELSRSPFPTHHLAQ	1200
QY	1201	CYRRAKKLESSEENULSDEDELPCFOHLLFEKVNNIPSOSTRHSTVATECLSKNTENL	1260
Db	1201	CYRRAKKLESSEENULSDEDELPCFOHLLFEKVNNIPSOSTRHSTVATECLSKNTENL	1260
QY	1261	LSLKNSLINDCSNOVILAKASOEHHLISEETKCSASFSSQCSLEBDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNSLINDCSNOVILAKASOEHHLISEETKCSASFSSQCSLEBDLTANTNTQDPFLIGS	1320
QY	1321	SKOMRHOSEBQCVGLSDKELVSDDERGTGLEENNOEBOQSDMSNLGEAASGCESETSVSE	1380
Db	1321	SKOMRHOSEBQCVGLSDKELVSDDERGTGLEENNOEBOQSDMSNLGEAASGCESETSVSE	1380
QY	1381	DCSGSSQSDILITTOQORDTMOHNLLKLOEAMALEFAYLEBOHSGOPSNSYPTIISOSALE	1440
Db	1381	DCSGSSQSDILITTOQORDTMOHNLLKLOEAMALEFAYLEBOHSGOPSNSYPTIISOSALE	1440
QY	1441	DLRNEQOSTSEKAVLTJSQKSSEYPIQSONPEGLISADKFEVSADSDSTSKNKEBGEVERSSPSK	1500
Db	1441	DLRNEQOSTSEKAVLTJSQKSSEYPIQSONPEGLISADKFEVSADSDSTSKNKEBGEVERSSPSK	1500
QY	1501	CPSLDDRWYMHSCSGLONRNPNSOEELIKVYDVEEQOLEBSGPHDLTETVLPQODLEG	1560
Db	1501	CPSLDDRWYMHSCSGLONRNPNSOEELIKVYDVEEQOLEBSGPHDLTETVLPQODLEG	1560
QY	1561	TPYLESGISLSDPESDPSEDRAPEASARVGNIPSTJALKYPOKJVAESASQSPAATTT	1620
Db	1561	TPYLESGISLSDPESDPSEDRAPEASARVGNIPSTJALKYPOKJVAESASQSPAATTT	1620
QY	1621	DTAGYNAMEESVSRKPELTASTERVNKRMSNAYVGLTPEEFMLVYFKARKHHTLTNLJ	1680
Db	1621	DTAGYNAMEESVSRKPELTASTERVNKRMSNAYVGLTPEEFMLVYFKARKHHTLTNLJ	1680

Qy	1661	TEETTHVYMKDAEYVCERTLKYTLGLAGGKWWVSYFMVQSIKERKLNHDEYRGDV	1740
Dd	1681	TEETTHVYMKDAEYVCERTLKYTLGLAGGKWWVSYFMVQSIKERKLNHDEYRGDV	1740
Qy	1741	VNGNHHGPKRARESDRKIFRGLEICCYGPTNMPDLEMMVQLGASVYKLSFTL	1800
Dd	1741	VNGNHHGPKRARESDRKIFRGLEICCYGPTNMPDLEMMVQLGASVYKLSFTL	1800
Qy	1801	GTGVHPYIVVQPDAMTEDNGFHAIGQKCEAPVYTRREWLDSVALYQCELDITYLIPQIPH	1860
Dd	1801	GTGVHPYIVVQPDAMTEDNGFHAIGQKCEAPVYTRREWLDSVALYQCELDITYLIPQIPH	1860
Qy	1861	SHY 1863	
Dd	1861	SHY 1863	

RESULT 8  
ABG01696  
ID ABG01696 standard; Protein, 2353 AA.  
XX  
AC ABG01696;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #1687.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dermanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS65883.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PS  
PS Claim 20: SEQ ID No 32055; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

SQ Sequence 2353 AA;

Query Match	99.98;	Score 9642;	DB 22;	Length 2353;
-------------	--------	-------------	--------	--------------

Best Local Similarity 99.98; Pred. No. 0;  
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSATREVEQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKPSQ 60

Db 491 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMKLLNQKKGPSQ 550

QY 61 CPICKNDITKRSLOESTRFSQVLEELKICAFQIDTGLEANSYNFAKKENNSPEHLKD 120

Db 551 CPICKNDITKRSLOESTRFSOLVEELKIIICAFOLDTGLEYSYNSYFAKKENNSPEHLKD 610

QY 121 EVSIIQSMGYRRAKRLLOSEPENPSLOETSLSVQLSNGTVRTLRTKORIOPOKTSVYI 180

Db 611 EVSIIQSMGYRNRAKRLQSEPENPSLQETSLSVQLSNGTVRTLRTKQRIQPOKTSYI 670

181 ELGSDSEDVKNATYCSVGDELLQITPQGRDEISLDSAKKAACEFSETDVTNTEHHQ 240

Db 671 ELGSDSSEDIVNKATYCSVGDELLQITPGTRDEISLDSAKKACEFSETDVTNTEHHQ 730

QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTHASSLQHENSLLTKDRMVE 300

Db 731 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTINTHASSLQHENSLLTKDRMVE 790

QY 301 KAEFCNKSQPGLARSHNRWAGSKETCNDRTPTSTEKKVDLNADPLCERKEWNKQKLP 360

Db 791 KAEECNKSQPLARSQHNRWAGSKETCNDRTPTSTEEKVDLNADPLCERKEWNKQKLP 850

QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELLGSDSDSHDGESESNNAKVADVLVNEVD 420

Db 851 SENPRDTEVPWITLNSSIQKVNEMFSRSEDELLGSDSDSHDGESESESNKAVADVLVDLNEVD 910

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKA\$LPNLSHVTEN 480

Db 911 EYSGSSEKIDLLASDPHEALICKSERVHSKSVEESNIEDKIFGKTYRKKASLPNLSHVTE 970

QY 481 LIIGAFVTEPQIIQERPLTNKLKRRRPTSGLHPEDFIKKADLAVQKTPMINQGTNQTE 540

Db 971 LIIGAVTEPQIIQERPLTNKLKRRKRPSTGLHPEDEFIKKADLAVQKTPMINQGTNQT 1030

541 QNGVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600

Db 1031 QNGVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISMELENI 1090

QY 601 HNSKAPKKNRLRRKSSTRIHALELVSRNLSPPNCTELQIDSCSSSEIKKKKYNQMPV 660

Db 1091 HNSKAPKKNR.LRRKSSTRHIALELVSRNLSPNCTELQIDSCSSSEIKKKKKYNQMPV 1150

661 R HSRNLQ LMEGKEPATGAKKSNKPNEQT SKRHSDTFPELKL TNAPGSFTKCSNTSELKE 720

Db 1151 RHRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDTFPELKLTNAPGSFTKCSNTSELKE 1210

721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLQTERSVESSISLVPGTGYCQ 780

Db 1211 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLTQERSVSSSSISLVPGTDTCTQ 1270

QY 781 ESISLLEVSTLGAKTEPNKCVSQCFAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNHS 840

Db 1271 ESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 1330

QY 841 RETSIEMEESELDAYLONTFKVSKRQSFALFSPGNAEEECATFSAHSGSLKKQSPVT 900

Db 1331 RETSIEMESELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPVT 1390

QY 901 FECEÖKEENÖGKNESNIKPVÖTVNITAGFPVVGÖKDKPVDNAKCSIKGGSRECLSSÖFRG 960

Db 1391 FECEQKEENQGNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRCFLSSQFRG 1450



```

QY 961 NETGLTPKKGHLQNPYRNPPLPPIKSEYKTKCKNNLEENFEHSMSPEREMENIP 1020
DB 1451 NETGLTPKKGHLQNPYRNPPLPPIKSEYKTKCKNNLEENFEHSMSPEREMENIP 1510
QY 1021 STVSTSRNNIRENVEKKAASSSSININVGSSSTNEVGSSINIGSSDENICQALGRNGPKL 1080
DB 1511 STVSTSRNNIRENVEKKAASSSSININVGSSSTNEVGSSINIGSSDENICQALGRNGPKL 1570
QY 1081 NAMLRGLVQLPEVYKOSLPGSNCKHPEIKKOEYEEVQVNTDPSFYLSDNLEOPMGSS 1140
DB 1571 NAMLRGLVQLPEVYKOSLPGSNCKHPEIKKOEYEEVQVNTDPSFYLSDNLEOPMGSS 1630
QY 1141 HASQVCSPTDLDLDDGETKEDTSFAENDIKESSAVFSKSVOKGELSRSPSPPTHTHLAQ 1200
DB 1631 HASQVCSPTDLDLDDGETKEDTSFAENDIKESSAVFSKSVOKGELSRSPSPPTHTHLAQ 1690
QY 1201 GYRGAACKLSESEENINSEDEELPCFOHLLFGKVNNPISOSTRHSVATRECLSKNNEENL 1260
DB 1691 GYRGAACKLSESEENINSEDEELPCFOHLLFGKVNNPISOSTRHSVATRECLSKNNEENL 1750
QY 1261 LSLKNSLNDCSNQVILAKASOEHLSEETKCSASLFSOSCELEDLTANTNQDPFLIGS 1320
DB 1751 LSLKNSLNDCSNQVILAKASOEHLSEETKCSASLFSOSCELEDLTANTNQDPFLIGS 1810
QY 1321 SKQMHQSSQGVGLSDKELVSDDERGTGLEBNNOEODSMDNLGEAASGCESETSVSE 1380
DB 1811 SKQMHQSSQGVGLSDKELVSDDERGTGLEBNNOEODSMDNLGEAASGCESETSVSE 1870
QY 1381 DCSGLSSQSDILTTQORDTMOHNLKLOEAMAELEVLFOHSGOPSNSVPTITSSSALE 1440
DB 1871 DCSGLSSQSDILTTQORDTMOHNLKLOEAMAELEVLFOHSGOPSNSVPTITSSSALE 1930
QY 1441 DLRNEQOSTSEKAVLTSOKSESEYPISQNEPGLSADKFEVSADSTSKNNEPEYERSPPSK 1500
DB 1931 DLRNEQOSTSEKAVLTSOKSESEYPISQNEPGLSADKFEVSADSTSKNNEPEYERSPPSK 1990
QY 1501 CPSLDDRWYMHSCSSGLQNRNTPSOBELIKVVDVEEQULEESGPHDITSTLPRODLEG 1560
DB 1991 CPSLDDRWYMHSCSSGLQNRNTPSOBELIKVVDVEEQULEESGPHDITSTLPRODLEG 2050
QY 1561 TPYLESGISLFSDDPESDPESDRAPESARVGNIPSTSLAKVPOLKVAESASPAAHNT 1620
DB 2051 TPYLESGISLFSDDPESDPESDRAPESARVGNIPSTSLAKVPOLKVAESASPAAHNT 2110
QY 1621 DTAGYNAMEESVSRKPELTASTERYNKRMSWVSGLAPEEFMLVYKFKARKHITLTNLI 1680
DB 2111 DTAGYNAMEESVSRKPELTASTERYNKRMSWVSGLAPEEFMLVYKFKARKHITLTNLI 2170
QY 1681 TEETHHYVAKTDAEFVCEERTLKYPFLGIAGKMWVSYFWVTOSIKERKMLNEHDFEYRGDV 1740
DB 2171 TEETHHYVAKTDAEFVCEERTLKYPFLGIAGKMWVSYFWVTOSIKERKMLNEHDFEYRGDV 2230
QY 1741 VNGRNHOGPKRARESDOKRIFRGLEICCYGPTNNPQDLEMMVOLLGASVYKELSSFTL 1800
DB 2231 VNGRNHOGPKRARESDOKRIFRGLEICCYGPTNNPQDLEMMVOLLGASVYKELSSFTL 2290
QY 1801 GTGVHPYVQPPDAMTENGFAIGQMCAPVYTRRWLDSVALYQCELDYLLIPQIPH 1860
DB 2291 GTGVHPYVQPPDAMTENGFAIGQMCAPVYTRRWLDSVALYQCELDYLLIPQIPH 2350
QY 1861 SHY 1863
DB 2351 SHY 2353

```

RESULT 9  
AAR81490 standard: Protein: 1863 AA.

ID AAR81490:  
AC AAR81490:  
XX  
DT 01-OCT-1996 (first entry)

```

XX XX BRCA1 mutant from patient 98.
DE XX
XX XX
XX XX Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;
KM KM antibody production; germline alteration; probe; lesion neoplasia; human;
KW KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX OS
XX XX Homo sapiens.
XX XX
XX XX Key Location/Qualifiers
FT FT Misc-difference 271 /note="V271M"
XX XX
XX XX W09605306-A2.
XX XX
XX XX 22-FEB-1996.
XX XX
XX XX 11-AUG-1995: 95WO-US10202.
XX XX
XX XX 07-JUN-1995: 95US-0483553.
XX XX 12-AUG-1994: 94US-0289221.
XX XX 02-SEP-1994: 94US-0300266.
XX XX 16-SEP-1994: 94US-0308104.
XX XX 29-NOV-1994: 94US-0348824.
XX XX 24-MAR-1995: 94US-0409305.
XX XX 07-JUN-1995: 95US-0480784.
XX XX
XX XX (MYRI-) MYRIAD GENETICS INC.
XX XX (CANC-) CANCER INST.
XX XX (RECH-) CENT RECH DU CHUL.
XX XX
XX XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX XX WPI: 1996-139702/14.
XX XX N-PSDB: AAT17446.
XX XX
XX XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX XX PT gene - for diagnosis and therapy of human breast and ovarian cancer
XX XX PS and for diagnosing pre-disposition to these cancers
XX XX
XX XX Claim 1: 218bp; English.
XX XX
XX XX AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
XX XX encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
XX XX (see AAR81481 for wild type protein). These mutations can be used as
XX XX immunogens for antibody production. The mutant BRCA1 genes encoding
XX XX these sequences have at least 1 mutation or polymorphism in comparison
XX XX to the wild type cDNA (see AAT17438 for wild type). By detecting a
XX XX germline alteration in the wild type BRCA1 gene, a predisposition for
XX XX breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
XX XX isolated from a tissue sample from a subject has a probe, corresponding
XX XX to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
XX XX allele-specific probe for a mutation of it), added to it. The conditions
XX XX allow for hybridisation of the probe to the mRNA, and any hybridisation
XX XX which occurs is detected. Alternatively the BRCA1 gene in the tissue
XX XX sample is isolated, and a shift in electrophoretic mobility of single
XX XX stranded DNA from the sample on a non-denaturing polyacrylamide gel
XX XX indicates a mutation. These methods of detection can also diagnose a
XX XX lesion neoplasia associated with the BRCA1 locus. The methods may be
XX XX used in gene therapy, protein replacement therapy and protein mimetics,
XX XX and may be used to screen for drugs in cancer therapy.
XX XX
XX XX Sequence 1863 AA:

```

Query Match 99.9%; Score 9639; DB 17; Length 1863;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MDLSALRYEEVQYNNINAMOKITLPCICLLEIKPEVSTKCHIFPCFCKLNLQKKGPSP 60
DB 1 MDLSALRYEEVQYNNINAMOKITLPCICLLEIKPEVSTKCHIFPCFCKLNLQKKGPSP 60
QY 61 CPLCKNDITKRSLOESTRFSQLVEELIKITICAFQDGTGEYANSYNFAKKENNSPEHLKD 120

```



```

|||||
Db 61 CPLCKNDITKRSLOESTRFSQLVTEBELLKIIAFOLDITGLEVANSYNFAKKNENSPHKLKD 120
Qy 121 EVSTIIOSMGYNRAKRLLOSEPNSLOETISLSVOLNLGTVRLTRKORLOPOKTSYVI 180
Db 121 EVSTIIOSMGYNRAKRLLOSEPNSLOETISLSVOLNLGTVRLTRKORLOPOKTSYVI 180
Qy 181 ELGSDSSEDPYNAKATYCSVGDQELLQITPOGTREISLSAKKACESESDVTNTEHHQ 240
Db 181 ELGSDSSEDPYNAKATYCSVGDQELLQITPOGTREISLSAKKACESESDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVEPCGTNTHASLOHENSLLITKDRNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVEPCGTNTHASLOHENSLLITKDRNVE 300
Qy 301 KAEFCNKSQOPGLARSOHNWAGSKETCNDRTRTSTKTKVDLNDPLCEKREKMKOKLPC 360
Db 301 KAEFCNKSQOPGLARSOHNWAGSKETCNDRTRTSTKTKVDLNDPLCEKREKMKOKLPC 360
Qy 361 SENRDTEDYPMITLNGSIOKVNMFPSSDELGSDSDHGESESNKAVADVLVLEVD 420
Db 361 SENRDTEDYPMITLNGSIOKVNMFPSSDELGSDSDHGESESNKAVADVLVLEVD 420
Qy 421 EYSGSSSEKIDLLASDPHEALICKSERVHSKSVESENIEDKIFGKTYRRKASLPLNLSHTEN 480
Db 421 EYSGSSSEKIDLLASDPHEALICKSERVHSKSVESENIEDKIFGKTYRRKASLPLNLSHTEN 480
Qy 481 LIIAFVTEPOIIOERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINGTQOTE 540
Db 481 LIIAFVTEPOIIOERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINGTQOTE 540
Qy 541 QNGGVMTNNGSHENKTKGDSIONEKNPNPIESLEKESAFKTKAEPISSTISNNELEINI 600
Db 541 QNGGVMTNNGSHENKTKGDSIONEKNPNPIESLEKESAFKTKAEPISSTISNNELEINI 600
Qy 601 HNSKAPKNRLRRKSTRHIALELVYSRNLSPNCTELOIDCSSSEELKKKYKNOMPV 660
Db 601 HNSKAPKNRLRRKSTRHIALELVYSRNLSPNCTELOIDCSSSEELKKKYKNOMPV 660
Qy 661 RHSRNILOMEGKEPATGAKKSNKNEQTSKRHSDTPELKLITNAPSPFTKCSMTSELKE 720
Db 661 RHSRNILOMEGKEPATGAKKSNKNEQTSKRHSDTPELKLITNAPSPFTKCSMTSELKE 720
Qy 721 FVNSPLPREKEEKELETVKVSNNNAEDPKDMLSGERVLQTERSVESSTISLVPCTDVGTO 780
Db 721 FVNSPLPREKEEKELETVKVSNNNAEDPKDMLSGERVLQTERSVESSTISLVPCTDVGTO 780
Qy 781 ESISLLEVSTLGAKTEPNKCVSQAFAENPKGLIHGCSKDNRDTEGFKYPLIGHEVNS 840
Db 781 ESISLLEVSTLGAKTEPNKCVSQAFAENPKGLIHGCSKDNRDTEGFKYPLIGHEVNS 840
Qy 841 RETSIEMEESELDQOYONTFKYSKROSPALFSNPGNAEECAFTSAHSGSLKQSPKVT 900
Db 841 RETSIEMEESELDQOYONTFKYSKROSPALFSNPGNAEECAFTSAHSGSLKQSPKVT 900
Qy 901 FECQKEENOGKNBSNIKPVQTVNITAGFPVYGOKDPVDNAKCSIGGSFECCLSSQPRG 960
Db 901 FECQKEENOGKNBSNIKPVQTVNITAGFPVYGOKDPVDNAKCSIGGSFECCLSSQPRG 960
Qy 961 NETGLITPNKHGILLONPYRIPLPIKSFYKTKCKNULNEENFEHSMSPREKNGENIP 1020
Db 961 NETGLITPNKHGILLONPYRIPLPIKSFYKTKCKNULNEENFEHSMSPREKNGENIP 1020
Qy 961 NETGLITPNKHGILLONPYRIPLPIKSFYKTKCKNULNEENFEHSMSPREKNGENIP 1020
Db 961 NETGLITPNKHGILLONPYRIPLPIKSFYKTKCKNULNEENFEHSMSPREKNGENIP 1020
Qy 1021 STVSTIRNNIRRENVFEASSNINEVGSSINELGSSDENIOAELGNRGPKL 1080
Db 1021 STVSTIRNNIRRENVFEASSNINEVGSSINELGSSDENIOAELGNRGPKL 1080
Qy 1081 NAMRLGLVLOPEVYKOLPGSNCKHPEIKOEYEEVQVTVTDSPLISNLEOPQSS 1140
Db 1081 NAMRLGLVLOPEVYKOLPGSNCKHPEIKOEYEEVQVTVTDSPLISNLEOPQSS 1140
Qy 1141 HASQVCEPDDLLDDEIEIKEDTISFAENDIKESSAVFSKSVQKELSRSPFPTHHLAQ 1200
Db 1141 HASQVCEPDDLLDDEIEIKEDTISFAENDIKESSAVFSKSVQKELSRSPFPTHHLAQ 1200

```

```

Db 1141 HASQVCEPDDLLDDEIEIKEDTISFAENDIKESSAVFSKSVQKELSRSPFPTHHLAQ 1200
Qy 1201 GYRRGAKKLESSSEENISSDEBELPCFOHLLFGKNNNIPSOSTRSTYATBELSKNTEENL 1260
Db 1201 GYRRGAKKLESSSEENISSDEBELPCFOHLLFGKNNNIPSOSTRSTYATBELSKNTEENL 1260
Qy 1261 LSLKNSLNDCSNOYIILAKASOEHLSEETKCSASLFSQCELEDLTANTTOOPFLIGS 1320
Db 1261 LSLKNSLNDCSNOYIILAKASOEHLSEETKCSASLFSQCELEDLTANTTOOPFLIGS 1320
Qy 1321 SKQMRHQSSEOGVGLSKELVSDDERGTGLEENNOEBSMDSNLGEAASCESETSVSE 1380
Db 1321 SKQMRHQSSEOGVGLSKELVSDDERGTGLEENNOEBSMDSNLGEAASCESETSVSE 1380
Qy 1381 DCSGLSSQSDLLTTOQNDTMOHNLIKIQOEAELEAVLEQSGOPSNSYPSIISDSALE 1440
Db 1381 DCSGLSSQSDLLTTOQNDTMOHNLIKIQOEAELEAVLEQSGOPSNSYPSIISDSALE 1440
Qy 1441 DLRRPEOSTSEKAAVLTSQSSSEYPISONPEGLSADKFEVSADSSSTSKNKEGVERSSPSK 1500
Db 1441 DLRRPEOSTSEKAAVLTSQSSSEYPISONPEGLSADKFEVSADSSSTSKNKEGVERSSPSK 1500
Qy 1501 CPSLDDRMWYHSCSGSLQNRNYPQOEELIKVYVDEEQOLEESGPHDLTETSYLEPRODLEG 1560
Db 1501 CPSLDDRMWYHSCSGSLQNRNYPQOEELIKVYVDEEQOLEESGPHDLTETSYLEPRODLEG 1560
Qy 1561 TPYLESGISLFSDDPESDPSDDRAPESARVGNIPSSSALKVPOLKVAESQSPAAHATT 1620
Db 1561 TPYLESGISLFSDDPESDPSDDRAPESARVGNIPSSSALKVPOLKVAESQSPAAHATT 1620
Qy 1621 DTAGYNAMEEVSSEKPELFTASTERVNRKMSMNVYSGLTPEEFMLVYFARKHHTLTNLI 1680
Db 1621 DTAGYNAMEEVSSEKPELFTASTERVNRKMSMNVYSGLTPEEFMLVYFARKHHTLTNLI 1680
Qy 1681 TEETTHVVMKTDABEVCERTLKYFLGIAGCKMVSYTFWVTQSIKERMLNHEDEVRGDV 1740
Db 1681 TEETTHVVMKTDABEVCERTLKYFLGIAGCKMVSYTFWVTQSIKERMLNHEDEVRGDV 1740
Qy 1741 VNGRHHGQPKRAROSORRTFRGLFTICCYGFTWMPDOLLEMMVOLCGASVYKELSSFTL 1800
Db 1741 VNGRHHGQPKRAROSORRTFRGLFTICCYGFTWMPDOLLEMMVOLCGASVYKELSSFTL 1800
Qy 1801 GTGVHPIVVOVPDAMTEDENGFHAIGOMCEAPVTRRENVLDVALYOCOLDTYLIPQIPH 1860
Db 1801 GTGVHPIVVOVPDAMTEDENGFHAIGOMCEAPVTRRENVLDVALYOCOLDTYLIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 10
AAR81500
ID AAR81500 standard; Protein; 1863 AA.
XX
AC AAR81500;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample sets YN98 and YN7.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT Misc-difference 271
XX /note- "v271m"
XX
PN W09605306-A2.
XX

```



Db 1381 DCGSLSSQSDILTTQOQDMQHNLIKQOEMAELEAVLEQHGQSPSNYSIISSDALE 1440  
 QY 1441 DLNPEOSTSEKAVLSQKSEEPISQNPGLSADFEVADSTSKNKPGEVRSPPSK 1500  
 Db 1441 DLNPEOSTSEKAVLSQKSEEPISQNPGLSADFEVADSTSKNKPGEVRSPPSK 1500  
 QY 1501 CPSLIDRWYHSCSGSLQNNRNPQOELLKVVDVEEQOLESGPHDLTETSYLPRODLEG 1560  
 Db 1501 CPSLIDRWYHSCSGSLQNNRNPQOELLKVVDVEEQOLESGPHDLTETSYLPRODLEG 1560  
 QY 1561 TPPLSGISLFSDDPESDPEEDRAPASARVGNIPSSTSALKVPOLKYAESAGSPAAAHHT 1620  
 Db 1561 TPPLSGISLFSDDPESDPEEDRAPASARVGNIPSSTSALKVPOLKYAESAGSPAAAHHT 1620  
 QY 1621 DTGYNAMSESVREKPELTASTERYNKRMSVSGLTPEEMLVYKFAKHHTLTNL 1680  
 Db 1621 DTGYNAMSESVREKPELTASTERYNKRMSVSGLTPEEMLVYKFAKHHTLTNL 1680  
 QY 1681 TEETTHVYMKTDAEFYCEKTLKYLFIAGGKVVSYFWYQSIKERKMLNEHDFEVRGDV 1740  
 Db 1681 TEETTHVYMKTDAEFYCEKTLKYLFIAGGKVVSYFWYQSIKERKMLNEHDFEVRGDV 1740  
 QY 1741 VGNRNHOGPKRARESDOKIFRGLFICCGPFTNMPDLEMMVQLGASVKEISSFTL 1800  
 Db 1741 VGNRNHOGPKRARESDOKIFRGLFICCGPFTNMPDLEMMVQLGASVKEISSFTL 1800  
 QY 1801 GTGVHPVYVQPDAMTDENGFHAIGOMCEAPVYTRREWLVDSVALYCCQELDTYLLIPIQIRH 1860  
 Db 1801 GTGVHPVYVQPDAMTDENGFHAIGOMCEAPVYTRREWLVDSVALYCCQELDTYLLIPIQIRH 1860  
 QY 1861 SHY 1863  
 Db 1861 SHY 1863

RESULT 11  
 AAR81522 standard; Protein; 1863 AA.  
 AAR81522:  
 02-OCT-1996 (first entry)  
 BRCA1 mutant from sample set MSK12871.  
 Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;  
 antibody production; germline alteration; probe; lesion neoplasia; human;  
 gene therapy; protein replacement therapy; protein mimetic; BRCA1.  
 Homo sapiens.  
 Key Location/Qualifiers  
 Misc-difference 1219 /note="E1219D"  
 W09605306-A2.  
 22-FEB-1996.  
 11-AUG-1995: 95MO-US10202.  
 07-JUN-1995: 95US-0483553.  
 12-AUG-1994: 94US-0289221.  
 16-SEP-1994: 94US-0300266.  
 29-NOV-1994: 94US-0308104.  
 24-MAR-1995: 94US-0348824.  
 07-JUN-1995: 95US-0409305.  
 07-JUN-1995: 95US-0480784.  
 (MYRI-) MYRIAD GENETICS INC.  
 (CANC-) CANCER INST.  
 (RECH-) CENT RECH DU CHUL.

XX Durocher F, Eml M, Nakamura Y, Simard J, Shattuck-Bidens DM;  
 PI WPI: 1996-139702/14.  
 DR N-PSDB; AAT17479.  
 XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers  
 XX  
 PS Claim 1: ; 218pp: English.  
 XX  
 CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein  
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)  
 CC (see AAR81481 for wild type protein). These mutations can be used as  
 CC immunogens for antibody production. The mutant BRCA1 genes encoding  
 CC these sequences have at least 1 mutation or polymorphism in comparison  
 CC to the wild type cDNA (see AAT17438 for wild type). By detecting a  
 CC germline alteration in the wild type BRCA1 gene, a predisposition for  
 CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA  
 CC isolated from a tissue sample from a subject has a probe, corresponding  
 CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an  
 CC allele-specific probe for a mutation of it), added to it. The conditions  
 CC allow for hybridisation of the probe to the mRNA, and any hybridisation  
 CC which occurs is detected. Alternatively the BRCA1 gene in the tissue  
 CC sample is isolated, and a shift in electrophoretic mobility of single  
 CC stranded DNA from the sample on a non-denaturing polyacrylamide gel  
 CC indicates a mutation. These methods of detection can also diagnose a  
 CC lesion neoplasia associated with the BRCA1 locus. The methods may be  
 CC used in gene therapy, protein replacement therapy and protein mimetics,  
 CC and may be used to screen for drugs in cancer therapy.  
 CC  
 XX  
 SQ Sequence 1863 AA:  
 Query Match 99.9%; Score 9639; DB 17; Length 1863;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MDLSALRVEEYOVNINAMOKILECPICLFIKPEVSTKCHIFCKRCMLLNOKKGPSP 60  
 Db 1 MDLSALRVEEYOVNINAMOKILECPICLFIKPEVSTKCHIFCKRCMLLNOKKGPSP 60  
 QY 61 CPLCKNDITKRSLOESTRESQVLEELIKIICAFQDLTGLEYANSYNFAKKENNSPEHLND 120  
 Db 61 CPLCKNDITKRSLOESTRESQVLEELIKIICAFQDLTGLEYANSYNFAKKENNSPEHLND 120  
 QY 121 EVSIIOSMGYRNFAKLLQSEPNPSLOETSLVSQSLNIGVYTLTKRIQIOPKTSVYI 180  
 Db 121 EVSIIOSMGYRNFAKLLQSEPNPSLOETSLVSQSLNIGVYTLTKRIQIOPKTSVYI 180  
 QY 181 ELGSDSSEPTVYKATYCSVGDOELLOITPOGTFDELSLSAKKAACEFSEDTVTNEHHQ 240  
 Db 181 ELGSDSSEPTVYKATYCSVGDOELLOITPOGTFDELSLSAKKAACEFSEDTVTNEHHQ 240  
 QY 241 PSNNDLNTTEKRAERHPRKYQGSYSNLHVEFCGTNTASSLOHNSLSLLTKDRMAYE 300  
 Db 241 PSNNDLNTTEKRAERHPRKYQGSYSNLHVEFCGTNTASSLOHNSLSLLTKDRMAYE 300  
 QY 301 KAEFCNKSQOPGLARQHNRMAGSKETCDNRTPSTTEKVVLDNADPLCERKENNOKLPC 360  
 Db 301 KAEFCNKSQOPGLARQHNRMAGSKETCDNRTPSTTEKVVLDNADPLCERKENNOKLPC 360  
 QY 361 SENPRDTEVPMTLLSSIOKVNEMFSRSDDELGSDSDSHGSESNAAKVAADVLDVINEVD 420  
 Db 361 SENPRDTEVPMTLLSSIOKVNEMFSRSDDELGSDSDSHGSESNAAKVAADVLDVINEVD 420  
 QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSHSVESNIEDKIFPKTYRKKASLNLNLSHVEN 480  
 Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSHSVESNIEDKIFPKTYRKKASLNLNLSHVEN 480  
 QY 481 LITGAVTEPQIIOERPLTNLTKRRRPTSGLHPEDFIKADLAVOKTPEMINOCTNOTE 540  
 Db 481 LITGAVTEPQIIOERPLTNLTKRRRPTSGLHPEDFIKADLAVOKTPEMINOCTNOTE 540

```

QY 541 QNGOVNNTNSGHNKTKGDSIONENKPNPIESLEKESAFKTKAEPDISSISMMLELNI 600
    |||
Db 541 QNGOVNNTNSGHNKTKGDSIONENKPNPIESLEKESAFKTKAEPDISSISMMLELNI 600
QY 601 HNSKAPKRRRLRRKSSRIHHALELVYVSRNLSPNCTELOIDSCSSSEETKKKKYQMPV 660
    |||
Db 601 HNSKAPKRRRLRRKSSRIHHALELVYVSRNLSPNCTELOIDSCSSSEETKKKKYQMPV 660
QY 661 RRSRNLQMLEGKEPATGAKSKNPNQOTSRHSDPFPELKLINAPGSFTKCSNTSELKE 720
    |||
Db 661 RRSRNLQMLEGKEPATGAKSKNPNQOTSRHSDPFPELKLINAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERYLOTERSVSSSISLVGTDTYGTQ 780
    |||
Db 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERYLOTERSVSSSISLVGTDTYGTQ 780
QY 781 ESISLLEVSTLGAKTEPKKCVSQCAAFENPKGLHGCSKDNNDTEGFKYPLGHEVNH 840
    |||
Db 781 ESISLLEVSTLGAKTEPKKCVSQCAAFENPKGLHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAOYLONTFKYKROSFALFSPNGNAEPECATFSAHSGSLKKOSPKVT 900
    |||
Db 841 RETSIEMESELDAOYLONTFKYKROSFALFSPNGNAEPECATFSAHSGSLKKOSPKVT 900
QY 901 FECEOEKEENOGKNESNIKPVQTVNITAGFPVYQKDKPYDNACSIKGGSRFCLSSQFRG 960
    |||
Db 901 FECEOEKEENOGKNESNIKPVQTVNITAGFPVYQKDKPYDNACSIKGGSRFCLSSQFRG 960
QY 961 NETGLITPKKHGLONPYRIPLPLPIKSYKTKCKNMLEENEEHSMSPERMGNENLP 1020
    |||
Db 961 NETGLITPKKHGLONPYRIPLPLPIKSYKTKCKNMLEENEEHSMSPERMGNENLP 1020
QY 1021 STVSTISRNNIRENVEKASSNINEVGSTNEVGSSINEIGSSDNIOELGRNNGPKL 1080
    |||
Db 1021 STVSTISRNNIRENVEKASSNINEVGSTNEVGSSINEIGSSDNIOELGRNNGPKL 1080
QY 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKOEVEEYVQTVNDSFYLISDNLQPMGSS 1140
    |||
Db 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKOEVEEYVQTVNDSFYLISDNLQPMGSS 1140
QY 1141 HASOVCEETPDLLDGEIKEDTSFAENDIKESSAFESKSVQCGELSRSPPTHTHIAQ 1200
    |||
Db 1141 HASOVCEETPDLLDGEIKEDTSFAENDIKESSAFESKSVQCGELSRSPPTHTHIAQ 1200
QY 1201 GYRRGAKKLESSEENLSEDEELPCFOHLFGKVNNIPOSOSTHSTVATECLSKNTEENL 1260
    |||
Db 1201 GYRRGAKKLESSEENLSEDEELPCFOHLFGKVNNIPOSOSTHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILIAKASQEHHLSEETKCSASLEFSQCSLEEDLTJANTNDPPLIGS 1320
    |||
Db 1261 LSLKNSLNDCSNOVILIAKASQEHHLSEETKCSASLEFSQCSLEEDLTJANTNDPPLIGS 1320
QY 1321 SKQMRHOSQGVGLSDKELVSDDERGCTGLENNODEOSMSNLCEAASGCSESTSVSE 1380
    |||
Db 1321 SKQMRHOSQGVGLSDKELVSDDERGCTGLENNODEOSMSNLCEAASGCSESTSVSE 1380
QY 1381 DCSGSLSSOSDILITTOORDTMOHNLKLOQEMAELEVLFOHSGOPSNSYPSIISDSSALE 1440
    |||
Db 1381 DCSGSLSSOSDILITTOORDTMOHNLKLOQEMAELEVLFOHSGOPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEOSTSEKAVILTSOKSEYPISQNPEGLSADKFEVYSADSTSKNKEPVERSPSPK 1500
    |||
Db 1441 DLRNPEOSTSEKAVILTSOKSEYPISQNPEGLSADKFEVYSADSTSKNKEPVERSPSPK 1500
QY 1501 CPSLDDRWYHSCSGSLONKRNTPSOEELIKYVDVEEQOLEESPHDLTETSTYLPDRODLG 1560
    |||
Db 1501 CPSLDDRWYHSCSGSLONKRNTPSOEELIKYVDVEEQOLEESPHDLTETSTYLPDRODLG 1560
QY 1561 TPYLESGISLFSDDPESDPSEDRAPESARVGNIPSSTSALKVPOLKVAESAGSPAAAHHT 1620
    |||
Db 1561 TPYLESGISLFSDDPESDPSEDRAPESARVGNIPSSTSALKVPOLKVAESAGSPAAAHHT 1620

```

```

QY 1621 DTAGYNAMEESVREKPELTASTERYNKRMSVSGLTPEEFMLYKFAKHHITLNL 1680
    |||
Db 1621 DTAGYNAMEESVREKPELTASTERYNKRMSVSGLTPEEFMLYKFAKHHITLNL 1680
QY 1681 TEETHVAKNTDAEFCERTLKFFGIAGKVVSYFWVTOSIKRKKMLNHDPEVRGDV 1740
    |||
Db 1681 TEETHVAKNTDAEFCERTLKFFGIAGKVVSYFWVTOSIKRKKMLNHDPEVRGDV 1740
QY 1741 VNGRNHOGRRAESODRKIFRGLEICYGPTNMPDOLBMMVOLCGASVYKELSPFL 1800
    |||
Db 1741 VNGRNHOGRRAESODRKIFRGLEICYGPTNMPDOLBMMVOLCGASVYKELSPFL 1800
QY 1801 GTGVHPITVVQDPDAMEDNGFHAIGMCBAPVYTRBWLDSVALYQCOELDTYLIPI 1860
    |||
Db 1801 GTGVHPITVVQDPDAMEDNGFHAIGMCBAPVYTRBWLDSVALYQCOELDTYLIPI 1860
QY 1861 SHY 1863
    |||
Db 1861 SHY 1863

RESULT 12
AAR81536
ID AAR81536 standard; Protein; 1863 AA.
AC AAR81536;
DT 02-OCT-1996 (first entry)
DE BRCA1 mutant from PM07.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 1183
XX FT Misc-difference /note="K1183R"
XX
XX PN W09605306-A2.
XX
XX PD 22-FEB-1996.
XX
XX PF 11-AUG-1995; 95MO-US10202.
XX
XX PR 07-JUN-1995; 95US-0483553.
XX PR 12-AUG-1994; 94US-0289221.
XX PR 02-SEP-1994; 94US-0300266.
XX PR 16-SEP-1994; 94US-0308104.
XX PR 29-NOV-1994; 94US-0348824.
XX PR 24-MAR-1995; 95US-0409305.
XX PR 07-JUN-1995; 95US-0480784.
XX
XX PA (MYRI-) MYRIAD GENETICS INC.
XX PA (CANC-) CANCER INST.
XX PA (RECH-) CENT RECH DU CHUL.
XX
XX PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
XX DR WPI: 1996-139702/14.
XX DR N-PSDB: AAT17494.
XX
XX PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX PT gene - for diagnosis and therapy of human breast and ovarian cancer
XX PT and for diagnosing pre-disposition to these cancers
XX
XX PS Claim 1; : 218pp; English.
XX
XX CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
XX CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
XX CC (see AAR81481 for wild type protein). These mutations can be used as

```

CC immunogens for antibody production. The mutant BRCA1 genes encoding  
CC these sequences have at least 1 mutation or polymorphism in comparison  
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a  
CC germline alteration in the wild type BRCA1 gene, a predisposition for  
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA  
CC isolated from a tissue sample from a subject has a probe, corresponding  
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an  
CC allele-specific probe for a mutation of it), added to it. The conditions  
CC allow for hybridisation of the probe to the mRNA, and any hybridisation  
CC occurs is detected. Alternatively the BRCA1 gene in the tissue  
CC sample is isolated, and a shift in electrophoretic mobility of single  
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel  
CC indicates a mutation. These methods of detection can also diagnose a  
CC lesion neoplasia associated with the BRCA1 locus. The methods may be  
CC used in gene therapy, protein replacement therapy and protein mimetics,  
CC and may be used to screen for drugs in cancer therapy.

XX Sequence 1863 AA:

Query Match 99.9%; Score 9639; DB 17; Length 1863;

Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNIAMOKILECPICELIKEPVSTKCDHIFCKFCMKLLNQKKGPQ 60  
DB 1 MDLSALRVEEVQVNIAMOKILECPICELIKEPVSTKCDHIFCKFCMKLLNQKKGPQ 60  
QY CPLECKNDITRSLQESTRSQVLEELIKITICAFQDGLTEYANSYNPAKKENSPHKLND 120  
DB CPLECKNDITRSLQESTRSQVLEELIKITICAFQDGLTEYANSYNPAKKENSPHKLND 120  
QY 121 EVSIIOSMGYRNRAKRLQSEPEPNSLOETSLSVQSLNLTGVRTLTQKRIQPKTSVYI 180  
DB 121 EVSIIOSMGYRNRAKRLQSEPEPNSLOETSLSVQSLNLTGVRTLTQKRIQPKTSVYI 180  
QY 181 ELGSDSSEDTVNAKATCSVGDDELQITPOGTRDEISLDSAKKACEFSETDVTNEHHQ 240  
DB 181 ELGSDSSEDTVNAKATCSVGDDELQITPOGTRDEISLDSAKKACEFSETDVTNEHHQ 240  
QY 241 PSNNNDITTEKRAEHRPEKYOGSSVNLHVEPCGNTNTHASSIQHENSLLTKDKMANVE 300  
DB 241 PSNNNDITTEKRAEHRPEKYOGSSVNLHVEPCGNTNTHASSIQHENSLLTKDKMANVE 300  
QY 301 KAEFCNKSKOPGLARSOHNMAWAGSKETCNDRTPTSTEEKVLDNADPLCEKKEKNNKOLPC 360  
DB 301 KAEFCNKSKOPGLARSOHNMAWAGSKETCNDRTPTSTEEKVLDNADPLCEKKEKNNKOLPC 360  
QY 361 SENPRTEDVPWITLNSSTOKVNEWFSRSDDELIGSDSHDGESENAKAVADVLVNEVD 420  
DB 361 SENPRTEDVPWITLNSSTOKVNEWFSRSDDELIGSDSHDGESENAKAVADVLVNEVD 420  
QY 421 EYSGSEKIDLLASDHEALICKSERVHKSVEVSNIEDKIFGKTYRKKAISPLVSHVTEN 480  
DB 421 EYSGSEKIDLLASDHEALICKSERVHKSVEVSNIEDKIFGKTYRKKAISPLVSHVTEN 480  
QY 481 LIIGAFVTEPQIIQERPLTNLTKRRRPTSGLHPEDFIKKADLAVOKTEPMINQGTNOTE 540  
DB 481 LIIGAFVTEPQIIQERPLTNLTKRRRPTSGLHPEDFIKKADLAVOKTEPMINQGTNOTE 540  
QY 541 QNQVANNITNSGHEKTKGDSIONENPNPIESLEKESAFKAEPISSISIMMELNLI 600  
DB 541 QNQVANNITNSGHEKTKGDSIONENPNPIESLEKESAFKAEPISSISIMMELNLI 600  
QY 601 HNSKAKKRLRKRKSTRIHHALELVYVNRSLSPNCTELOIDSCSSEELKKKKYQOMPV 660  
DB 601 HNSKAKKRLRKRKSTRIHHALELVYVNRSLSPNCTELOIDSCSSEELKKKKYQOMPV 660  
QY 661 RHSRNLQLEMGKRPATGAKSKSNPNBOTSKRHSDPFPBLKLTNAGSFTKCSNTSELKE 720  
DB 661 RHSRNLQLEMGKRPATGAKSKSNPNBOTSKRHSDPFPBLKLTNAGSFTKCSNTSELKE 720  
QY 721 FVNPSPLPREKKEKLETYKVSNAEDPKDMLSGERYLOTERSVSSSISLVPGTDYGTQ 780  
DB 721 FVNPSPLPREKKEKLETYKVSNAEDPKDMLSGERYLOTERSVSSSISLVPGTDYGTQ 780

DB 721 FVNPSPLPREKKEKLETYKVSNAEDPKDMLSGERYLOTERSVSSSISLVPGTDYGTQ 780  
QY 781 ESTSILEVSTLGRKAKTEPNKCVSOCAAFENPKLJHCCSDNNNDPEGEFYPLGHEVNH 840  
DB 781 ESTSILEVSTLGRKAKTEPNKCVSOCAAFENPKLJHCCSDNNNDPEGEFYPLGHEVNH 840  
QY 841 RETSIEMEESELDQVLOANTFVKYSKRSQSFALFNNPGNAEECATFSAHSGSLKKQSPKV 900  
DB 841 RETSIEMEESELDQVLOANTFVKYSKRSQSFALFNNPGNAEECATFSAHSGSLKKQSPKV 900  
QY 901 FECEQKEENQGNKESNIPQVQVNTAGFPVVOQKRPVDNAKCSIKGSRFCLSSQFRG 960  
DB 901 FECEQKEENQGNKESNIPQVQVNTAGFPVVOQKRPVDNAKCSIKGSRFCLSSQFRG 960  
QY 961 NEGLGLTPNKHGLQNDPYRIRPLPIKSPKTCCKNLLLENEEHSMSPEREMGNTP 1020  
DB 961 NEGLGLTPNKHGLQNDPYRIRPLPIKSPKTCCKNLLLENEEHSMSPEREMGNTP 1020  
QY 1021 STVSTISRNRIRENVKREASSSNIINEVSSSTNEVGSSINEIGSSDENIQALGRNRGPKL 1080  
DB 1021 STVSTISRNRIRENVKREASSSNIINEVSSSTNEVGSSINEIGSSDENIQALGRNRGPKL 1080  
QY 1081 NAMLRGLVLOPEVYKOSLPESNCKHPEIKQOEYEVQVNTDPSPLYISDNLQEPKGS 1140  
DB 1081 NAMLRGLVLOPEVYKOSLPESNCKHPEIKQOEYEVQVNTDPSPLYISDNLQEPKGS 1140  
QY 1141 HASOVSEPPDDLDDGCEIKEDTSPFANDIKESAAFSKVQKGLSRSPPTHTHQA 1200  
DB 1141 HASOVSEPPDDLDDGCEIKEDTSPFANDIKESAAFSKVQKGLSRSPPTHTHQA 1200  
QY 1201 GYRGAKKLTSESEENLSEDEELPCFQHLIFGVNNIIPQSOTRHSVATPECLSKNTEENL 1260  
DB 1201 GYRGAKKLTSESEENLSEDEELPCFQHLIFGVNNIIPQSOTRHSVATPECLSKNTEENL 1260  
QY 1261 LSLKNSLNDCSNOVLIAKASQEHNLSEFTKCSALFSSQCELEDLTANTQDPLIGS 1320  
DB 1261 LSLKNSLNDCSNOVLIAKASQEHNLSEFTKCSALFSSQCELEDLTANTQDPLIGS 1320  
QY 1321 SKOMRHOSQOGVGLSDKELVSDDEERGTCLENNNOEDQMSDNLCEAASGCCSEFSVSC 1380  
DB 1321 SKOMRHOSQOGVGLSDKELVSDDEERGTCLENNNOEDQMSDNLCEAASGCCSEFSVSC 1380  
QY 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMALEAVLEQHGSPQNSVPSIISDSALE 1440  
DB 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOQEMALEAVLEQHGSPQNSVPSIISDSALE 1440  
QY 1441 DLNRPQOSTSEKAVLTQKSSSEYPISONPEGLSADKFEVSADOSTSKNKEPVERSSPSK 1500  
DB 1441 DLNRPQOSTSEKAVLTQKSSSEYPISONPEGLSADKFEVSADOSTSKNKEPVERSSPSK 1500  
QY 1501 CPSLDRRWVHSCSGSLQNNRNPVSOBELIKVYVVEEQOLEEGPRLUTETSYLPRODLG 1560  
DB 1501 CPSLDRRWVHSCSGSLQNNRNPVSOBELIKVYVVEEQOLEEGPRLUTETSYLPRODLG 1560  
QY 1561 TPYLESGISLFSDDPSPDSSEDRAPESARVGNIPSSSTALKVQLVAESAQSPAHAHTT 1620  
DB 1561 TPYLESGISLFSDDPSPDSSEDRAPESARVGNIPSSSTALKVQLVAESAQSPAHAHTT 1620  
QY 1621 DTAGYNAMEESVREKPELTAFTERVNRKMSVVSGLTPEEFMLVYKFAKHHITLTNLI 1680  
DB 1621 DTAGYNAMEESVREKPELTAFTERVNRKMSVVSGLTPEEFMLVYKFAKHHITLTNLI 1680  
QY 1681 TFEETTHVVKTAEEFCERTIKFGLIAGGKVVVSFWWTQOSTKERKMLNEHDFEVRGV 1740  
DB 1681 TFEETTHVVKTAEEFCERTIKFGLIAGGKVVVSFWWTQOSTKERKMLNEHDFEVRGV 1740  
QY 1741 VNGRNHOGKRAAESODRKIFRGLETCCGPTNMTDLEMMVQJCGASVYKELSSFTL 1800  
DB 1741 VNGRNHOGKRAAESODRKIFRGLETCCGPTNMTDLEMMVQJCGASVYKELSSFTL 1800  
QY 1801 GTGVNPIVVOVDPAMTLEDNGFHAIGOMCAPVYTRRWVLDVAALVQCELDITLYLIPQIPH 1860  
DB 1801 GTGVNPIVVOVDPAMTLEDNGFHAIGOMCAPVYTRRWVLDVAALVQCELDITLYLIPQIPH 1860

QY 1861 SHY 1863  
 |||  
 .Db 1861 SHY 1863

RESULT 13  
 AAR81540  
 ID AAR81540 standard; Protein; 1863 AA.  
 XX AAR81540;  
 AC AAR81540;  
 XX 02-OCT-1996 (first entry)  
 XX  
 XX BRCA1 mutant from PM20.  
 DE  
 XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;  
 KW antibody production; germline alteration; probe; lesion neoplasia; human;  
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1040 /note= "S1040N"  
 FT  
 XX MO9605306-A2.  
 XX  
 PD 22-FEB-1996.  
 XX  
 PD 11-AUG-1995; 95WO-US10202.  
 PF  
 XX 07-JUN-1995; 95US-0483553.  
 PR 12-AUG-1994; 94US-0289221.  
 PR 02-SEP-1994; 94US-0300266.  
 PR 16-SEP-1994; 94US-0308104.  
 PR 29-NOV-1994; 94US-0348824.  
 PR 24-MAR-1995; 95US-0409305.  
 PR 07-JUN-1995; 95US-0480784.  
 XX  
 PA (MYRI-) MYRIAD GENETICS INC.  
 PA (CANC-) CANCER INST.  
 PA (RECH-) CENT RECH DU CHUL.  
 XX  
 PI Durocher F, Eml M, Nakamura Y, Simard J, Shattuck-Eidens DM;  
 XX MPI: 1996-139702/14.  
 DR N-PSDB; AAT17502.  
 XX  
 PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers  
 XX  
 PS Claim 1; : 218pp; English.

CC and may be used to screen for drugs in cancer therapy.  
 XX  
 SQ Sequence 1863 AA;  
 Query Match 99.9%; Score 9639; DB 17; Length 1863;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSARVEEYGVNINAMOKILIECPICELIKPEPSTGCHIFKFCMLKLLNKKPSQ 60  
 |||  
 Db 1 MDLSARVEEYGVNINAMOKILIECPICELIKPEPSTGCHIFKFCMLKLLNKKPSQ 60

QY 61 CPLCKNDITKRSLOESTREFSOLVBEELKTIICAFOLDGLLEVANSYNFAKKENNSPEHLKD 120  
 |||  
 Db 61 CPLCKNDITKRSLOESTREFSOLVBEELKTIICAFOLDGLLEVANSYNFAKKENNSPEHLKD 120

QY 121 EVSIIISMGYRNRAKRLLOSEPENPSLOETSIVOLSMGLTVRLTKORIQPKTSVYI 180  
 |||  
 Db 121 EVSIIISMGYRNRAKRLLOSEPENPSLOETSIVOLSMGLTVRLTKORIQPKTSVYI 180

QY 181 ELGSDSEEDTVKATYCGVDOELQITPOGTBDEISLSAKKACESESTDVNTETHQ 240  
 |||  
 Db 181 ELGSDSEEDTVKATYCGVDOELQITPOGTBDEISLSAKKACESESTDVNTETHQ 240

QY 241 PSNNDLNTTEKRAAERHPEKYOGSSVNLHVEPCGTNTHASSLOHENSLLTKDRNVE 300  
 |||  
 Db 241 PSNNDLNTTEKRAAERHPEKYOGSSVNLHVEPCGTNTHASSLOHENSLLTKDRNVE 300

QY 301 KAEFCNKSKOPCLANSQNNRAGSKETCNDRTSTTEKKVVLNADPLCERKENMKOLPC 360  
 |||  
 Db 301 KAEFCNKSKOPCLANSQNNRAGSKETCNDRTSTTEKKVVLNADPLCERKENMKOLPC 360

QY 361 SENPRDTEVPWITLNNSTLOKYNEMFSRDELLSDSDSHDESSNAKVADVLVLEVD 420  
 |||  
 Db 361 SENPRDTEVPWITLNNSTLOKYNEMFSRDELLSDSDSHDESSNAKVADVLVLEVD 420

QY 421 EYSGSSEKIDLLASDPHALICKSERVHASKVESNEIDKIFGKTYRRKASLPNLSHTEN 480  
 |||  
 Db 421 EYSGSSEKIDLLASDPHALICKSERVHASKVESNEIDKIFGKTYRRKASLPNLSHTEN 480

QY 481 LIIGAFVEPQIIOGRPLTNLKKRRRPTSLGHEDEPIKADLAVOKTPEMINQNOTE 540  
 |||  
 Db 481 LIIGAFVEPQIIOGRPLTNLKKRRRPTSLGHEDEPIKADLAVOKTPEMINQNOTE 540

QY 541 ONGQVYNTTNSGHEKKTGDSIONENKNPNPESLEKESAFTRKAPLSSSTSNMELNI 600  
 |||  
 Db 541 ONGQVYNTTNSGHEKKTGDSIONENKNPNPESLEKESAFTRKAPLSSSTSNMELNI 600

QY 601 HNSKAPKKNRRLKRSSTRHIALLELVSRNLSPNCTELQIDSCSSSEETIKKKRYNOMPV 660  
 |||  
 Db 601 HNSKAPKKNRRLKRSSTRHIALLELVSRNLSPNCTELQIDSCSSSEETIKKKRYNOMPV 660

QY 661 RHSRNLQIMEGKEPATGAKKSNKPNEQTSKRHDSDTPEELKLTNAPSPFKCSWTSLKE 720  
 |||  
 Db 661 RHSRNLQIMEGKEPATGAKKSNKPNEQTSKRHDSDTPEELKLTNAPSPFKCSWTSLKE 720

QY 721 FVNPSLPREKEKEKETETKVSNNMEDPKDMLSEBRLVQTRSVSSSISILVPTDGTG 780  
 |||  
 Db 721 FVNPSLPREKEKEKETETKVSNNMEDPKDMLSEBRLVQTRSVSSSISILVPTDGTG 780

QY 781 ESISLLEVSTLGAKTEBPNKCVSOCAAFENRGLIHGCSKDNRRDTEGFFYPPLGHEVNH 840  
 |||  
 Db 781 ESISLLEVSTLGAKTEBPNKCVSOCAAFENRGLIHGCSKDNRRDTEGFFYPPLGHEVNH 840

QY 841 RETSIEMEESLDAQYLONTFKVSKROSFALEFNPNAEECATFSAHSSGLKQSKVY 900  
 |||  
 Db 841 RETSIEMEESLDAQYLONTFKVSKROSFALEFNPNAEECATFSAHSSGLKQSKVY 900

QY 901 FECEQKEENOKNENINIKPVQTVNTAGFPVYGOKDKPVDNAKSTIGSGRPLCSSQFRG 960  
 |||  
 Db 901 FECEQKEENOKNENINIKPVQTVNTAGFPVYGOKDKPVDNAKSTIGSGRPLCSSQFRG 960

QY 961 NETGLITNKHGLLONPRIPPLPPIKSFVTKCKKMLLENEFHSMSPERENGENITP 1020

```

Db      961 NEGLITPNKHGLQNPRIPLPEPIKSFVTKCKKMLLENEEHSMSEREGNGNIP 1020
Qy      1021 STVSTISRNNIRENVFEKASSNNINEYSGSTNEVGSSINEIGSSDEIQAEIGKRRPKL 1080
Db      1021 STVSTISRNNIRENVFEKASSNNINEYSGSTNEVGSSINEIGSSDEIQAEIGKRRPKL 1080
Qy      1081 NAMRLGLVLPQPEYVKQSLPGSNCKHPEIKKOEYEEVQVNTVDFSPYLSIDNEIQPMGSS 1140
Db      1081 NAMRLGLVLPQPEYVKQSLPGSNCKHPEIKKOEYEEVQVNTVDFSPYLSIDNEIQPMGSS 1140
Qy      1141 HASOVGSETPDDLLDQGEIKEDTSFANDIKESSAVFSKVQVQGLSRSPFTHTLAAQ 1200
Db      1141 HASOVGSETPDDLLDQGEIKEDTSFANDIKESSAVFSKVQVQGLSRSPFTHTLAAQ 1200
Qy      1201 GYRGAKKLESSEENLSEDEELPCFOHLLFGKVNNPISOSTRSHVATCLSKNTPENL 1260
Db      1201 GYRGAKKLESSEENLSEDEELPCFOHLLFGKVNNPISOSTRSHVATCLSKNTPENL 1260
Qy      1261 LSKNLSLNDCSNOVYLAKASQEHHLSEETKCSASLFSQCSLEDELTANTNTODPFLIGS 1320
Db      1261 LSKNLSLNDCSNOVYLAKASQEHHLSEETKCSASLFSQCSLEDELTANTNTODPFLIGS 1320
Qy      1321 SKQMRHOSQVGLSDKELVSDDEERKGTLEENNOEGQMDSNLGEAAGCSESTSVSE 1380
Db      1321 SKQMRHOSQVGLSDKELVSDDEERKGTLEENNOEGQMDSNLGEAAGCSESTSVSE 1380
Qy      1381 DCSGLSSQSDILFTQORDTQOHNLIKLOQMAELEAVLEQHGSPQSYSTIISDSSALE 1440
Db      1381 DCSGLSSQSDILFTQORDTQOHNLIKLOQMAELEAVLEQHGSPQSYSTIISDSSALE 1440
Qy      1441 DLBNPEQSTSEKAVLTSQKSESEYPISONPEGLSADKFEVADSSTSKNKEPVERSSPSK 1500
Db      1441 DLBNPEQSTSEKAVLTSQKSESEYPISONPEGLSADKFEVADSSTSKNKEPVERSSPSK 1500
Qy      1501 CPGLDQRMWYHSCSGSLQNNNYSQBELIKVYVVEEQQLSESGPHDLTETSYLPRDLEG 1560
Db      1501 CPGLDQRMWYHSCSGSLQNNNYSQBELIKVYVVEEQQLSESGPHDLTETSYLPRDLEG 1560
Qy      1561 TPYLESGISLFDSDPESDPESDRAPSARVGNIPSSTSALKVQOLKVAESAQSPAANTT 1620
Db      1561 TPYLESGISLFDSDPESDPESDRAPSARVGNIPSSTSALKVQOLKVAESAQSPAANTT 1620
Qy      1621 DTAGYNAMBEVSREKPELTASTERYNKRMVAVSGLTPEEFMLVYKFKARRHHITLTNLT 1680
Db      1621 DTAGYNAMBEVSREKPELTASTERYNKRMVAVSGLTPEEFMLVYKFKARRHHITLTNLT 1680
Qy      1681 TEETTHVVMKTDAEFCEKTLKFLGIAGKRWVSYFWVTQSIKERKMLNEHPREVGDV 1740
Db      1681 TEETTHVVMKTDAEFCEKTLKFLGIAGKRWVSYFWVTQSIKERKMLNEHPREVGDV 1740
Qy      1741 VGNRNHGPKRARESDRKTFRGLIEICCYGPTNMPDQLEMMVQOLGASVAVKELSEFTL 1800
Db      1741 VGNRNHGPKRARESDRKTFRGLIEICCYGPTNMPDQLEMMVQOLGASVAVKELSEFTL 1800
Qy      1801 GTGVNPIVYVQPDAMTEDNGFNALIGOMCEAPVYVTRVAVLDSVALYQCELDYTLIPQIP 1860
Db      1801 GTGVNPIVYVQPDAMTEDNGFNALIGOMCEAPVYVTRVAVLDSVALYQCELDYTLIPQIP 1860
Qy      1861 SHY 1863
Db      1861 SHY 1863

```

## RESULT 14

AAR81529 standard; Protein; 1863 AA.

```

ID AAR81529 standard; Protein; 1863 AA.
XX AAR81529;
AC AAR81529;
XX
XX 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample set MSR9646.

```

```

XX      Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;
KW      antibody production; germline alteration; probe; lesion neoplasia; human;
KW      gene therapy: protein replacement therapy; protein mimetic; BRCA1.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 1628
FT      "note" "M1628V"
XX
XX      W09605306-A2.
XX
XX      22-FEB-1996.
XX
XX      11-AUG-1995; 95WC-U010202.
XX
XX      07-JUN-1995; 95US-0483553.
XX      12-AUG-1994; 94US-0289221.
XX      02-SEP-1994; 94US-0300266.
XX      16-SEP-1994; 94US-0308104.
XX      29-NOV-1994; 94US-0348824.
XX      24-MAR-1995; 95US-0409305.
XX      07-JUN-1995; 95US-0480784.
XX
XX      (MYRI-) MYRIAD GENETICS INC.
XX      (CANC-) CANCER INST.
XX      (RECH-) CENT RECH DU CHUL.
XX
XX      Durocher F, Eml M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
XX      WPI; 1996-139702/14.
XX      N-PSDB; AAT17486.
XX
XX      New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX      PT gene for diagnosis and therapy of human breast and ovarian cancer
XX      PT and for diagnosing pre-disposition to these cancers
XX
XX      Claim 1; 218pp: English.
XX
XX      AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
XX      encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
XX      CC (see AAR81481 for wild type protein). These mutations can be used as
XX      CC immunogens for antibody production. The mutant BRCA1 genes encoding
XX      CC these sequences have at least 1 mutation or polymorphism in comparison
XX      CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
XX      CC germline alteration in the wild type BRCA1 gene, a predisposition for
XX      CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
XX      CC isolated from a tissue sample from a subject has a probe, corresponding
XX      CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
XX      CC allele-specific probe for a mutation of it), added to it. The conditions
XX      CC allow for hybridisation of the probe to the mRNA, and any hybridisation
XX      CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
XX      CC sample is isolated, and a shift in electrophoretic mobility of single
XX      CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
XX      CC indicates a mutation. These methods of detection can also diagnose a
XX      CC lesion neoplasia associated with the BRCA1 locus. The methods may be
XX      CC used in gene therapy, protein replacement therapy and protein mimetics,
XX      CC and may be used to screen for drugs in cancer therapy.
XX
XX      Sequence 1863 AA:
XX
XX      Query Match 99.9%; Score 9638; DB 17; Length 1863;
XX      Best Local Similarity 99.9%; Pred. No. 0;
XX      Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```



QY 121 EVSIISOMGYRRNRAKRLLOSEPENPSLOETSLSVOLSNLGTVRTLRKRIQIOPQKTSYVI 180  
 DB 122 EVSIISOMGYRRNRAKRLLOSEPENPSLOETSLSVOLSNLGTVRTLRKRIQIOPQKTSYVI 180  
 QY 181 EVGSDSEDTYKATYCSVGOEQLLOITPOGTIDEISLDSAKKACEFSEDTVTNTEHHQ 240  
 DB 181 EVGSDSEDTYKATYCSVGOEQLLOITPOGTIDEISLDSAKKACEFSEDTVTNTEHHQ 240  
 QY 241 PENNLNTEKRAAEHPEKRYOGSSYSNLAHEPCGNTNHAASLOHENSLLTLKDDMAYE 300  
 DB 241 PENNLNTEKRAAEHPEKRYOGSSYSNLAHEPCGNTNHAASLOHENSLLTLKDDMAYE 300  
 QY 301 KAEFCNKSQOPGLARQOHNRMAWSKETCNDRTPTSEKKYVDLNADELCEKRENNKOLPC 360  
 DB 301 KAEFCNKSQOPGLARQOHNRMAWSKETCNDRTPTSEKKYVDLNADELCEKRENNKOLPC 360  
 QY 361 SENPDTEDYVPIITLNNSTOKVNEWFSRSEDELIGSDSDHGESESNAKYADVLVLEND 420  
 DB 361 SENPDTEDYVPIITLNNSTOKVNEWFSRSEDELIGSDSDHGESESNAKYADVLVLEND 420  
 QY 421 EYSGSSEKIDLASDPEHALICKSERVHKSVESENTEDIKFGTKYRKASLPMLSHVTEN 480  
 DB 421 EYSGSSEKIDLASDPEHALICKSERVHKSVESENTEDIKFGTKYRKASLPMLSHVTEN 480  
 QY 481 LIIGAFTVEPOJIOERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTEPMINQOTNOTE 540  
 DB 481 LIIGAFTVEPOJIOERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTEPMINQOTNOTE 540  
 QY 541 ONGQVWNTNSGHEKTKDSTIONENPNPISLEKESAFKTAEPISISSIMBELNLT 600  
 DB 541 ONGQVWNTNSGHEKTKDSTIONENPNPISLEKESAFKTAEPISISSIMBELNLT 600  
 QY 601 HNSKAPKRRRLRRKSTRHIALELVVSNRNLSPNCTEJOIDSCSSSEETKRRKYQMPY 660  
 DB 601 HNSKAPKRRRLRRKSTRHIALELVVSNRNLSPNCTEJOIDSCSSSEETKRRKYQMPY 660  
 QY 661 RHRNRLQIMEGKEPATGAKKSNKPNEQTSKRHSDSTPELKLITNAGSFTKCSNTSELKE 720  
 DB 661 RHRNRLQIMEGKEPATGAKKSNKPNEQTSKRHSDSTPELKLITNAGSFTKCSNTSELKE 720  
 QY 721 FVNPSLPREKEKEKLETVKVSNNAPEDPKDMLSGEVLQTERSVESSTSLVPGTGYGQ 780  
 DB 721 FVNPSLPREKEKEKLETVKVSNNAPEDPKDMLSGEVLQTERSVESSTSLVPGTGYGQ 780  
 QY 781 ESISLLEVSTLGAKEPKNKCVSQAFAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNS 840  
 DB 781 ESISLLEVSTLGAKEPKNKCVSQAFAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNS 840  
 QY 841 RETSIEMESELDAQYLONTFKVYSKROSFALFSPNGNAEECATFSAHGSLKKQSPKVT 900  
 DB 841 RETSIEMESELDAQYLONTFKVYSKROSFALFSPNGNAEECATFSAHGSLKKQSPKVT 900  
 QY 901 FECBOKKEENOGKNEKNIKPVOTVNTAGPVYQKXKPYDNKCKSTKKGSRCLSSQFRG 960  
 DB 901 FECBOKKEENOGKNEKNIKPVOTVNTAGPVYQKXKPYDNKCKSTKKGSRCLSSQFRG 960  
 QY 961 NETGLITPNKHGLONPYRIPLFPKISFVKTKCKKNLLEENFEHSMSPEREMENP 1020  
 DB 961 NETGLITPNKHGLONPYRIPLFPKISFVKTKCKKNLLEENFEHSMSPEREMENP 1020  
 QY 1021 STVSTISRRNNTRENVFKKASSNINEVGSSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080  
 DB 1021 STVSTISRRNNTRENVFKKASSNINEVGSSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080  
 QY 1081 NAMRLGLVLOPEVYQSLPGSNCKHPETIKKOYEYEVYQVNVNDESYLLSDNLEQMGSS 1140  
 DB 1081 NAMRLGLVLOPEVYQSLPGSNCKHPETIKKOYEYEVYQVNVNDESYLLSDNLEQMGSS 1140  
 QY 1141 HASQVCESTPDDLDDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200  
 DB 1141 HASQVCESTPDDLDDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSPFTHTHLAQ 1200

QY 1201 GYRGAKKLESSEENLSEDEELPCFOHLLFGKVVNNIPSOSTRHSTVATECLSKNTEENL 1260  
 DB 1201 GYRGAKKLESSEENLSEDEELPCFOHLLFGKVVNNIPSOSTRHSTVATECLSKNTEENL 1260  
 QY 1261 LSLKNSLNDCSNOVILAKASOEHHUSEETKGCASLFPSSOCSELEDTANTNTODPFLIGS 1320  
 DB 1261 LSLKNSLNDCSNOVILAKASOEHHUSEETKGCASLFPSSOCSELEDTANTNTODPFLIGS 1320  
 QY 1321 SKOMRHOSQOGSLDKELVSDDERGTGLEENNOEBSMNSNGEASGCESTVSSE 1380  
 DB 1321 SKOMRHOSQOGSLDKELVSDDERGTGLEENNOEBSMNSNGEASGCESTVSSE 1380  
 QY 1381 DCSGSSOSDILITTOORDTMQHNLIKLOEMAELAVLEOHGSOPSNSYPSIISSSALE 1440  
 DB 1381 DCSGSSOSDILITTOORDTMQHNLIKLOEMAELAVLEOHGSOPSNSYPSIISSSALE 1440  
 QY 1441 DLNRPQOSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSADSTSKKKEGVERSSPSK 1500  
 DB 1441 DLNRPQOSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSADSTSKKKEGVERSSPSK 1500  
 QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVYDVEEQULESGPHDLTETSYLPRODLEG 1560  
 DB 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVYDVEEQULESGPHDLTETSYLPRODLEG 1560  
 QY 1561 TPYLESGISLFSDDPESDPSEDRAESARVNIIPSTSAKLVQPKVAESAQSPAATHT 1620  
 DB 1561 TPYLESGISLFSDDPESDPSEDRAESARVNIIPSTSAKLVQPKVAESAQSPAATHT 1620  
 QY 1621 DTAGYNAAMEESVSRKPELTASTERYNKRMSNVVGLTPEEMLYKRAKHHTLTNLI 1680  
 DB 1621 DTAGYNAAMEESVSRKPELTASTERYNKRMSNVVGLTPEEMLYKRAKHHTLTNLI 1680  
 QY 1681 TEETTHVAKTDAEFECERTLKYFLGIAGKVVVSYFWVTOSIKRKMILNHDDEVGRDY 1740  
 DB 1681 TEETTHVAKTDAEFECERTLKYFLGIAGKVVVSYFWVTOSIKRKMILNHDDEVGRDY 1740  
 QY 1741 VNGRNHQGPKRARESDRKIFRGLIICYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800  
 DB 1741 VNGRNHQGPKRARESDRKIFRGLIICYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800  
 QY 1801 GTGVHPPIVVOPDAMTEONGFHAIGOMCEAPVYRFEWLDVALYQCOELDTYLPOLPH 1860  
 DB 1801 GTGVHPPIVVOPDAMTEONGFHAIGOMCEAPVYRFEWLDVALYQCOELDTYLPOLPH 1860  
 QY 1861 SHY 1863  
 DB 1861 SHY 1863  
 RESULT 15  
 AAR81532  
 ID AAR81532 standard; Protein; 1863 AA.  
 XX  
 AC AAR81532;  
 DT 02-OCT-1996 (first entry)  
 XX  
 DE BRCA1 mutant from sample set MSK7542.  
 XX  
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;  
 KW antibody production; germline alteration; probe; lesion neoplasia; human;  
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.  
 XX  
 XX Homo sapiens.  
 XX  
 FH Key location/Qualifiers  
 FT Misc-difference 1852 /note= "T1852S"  
 XX  
 XX MO9605306-A2.  
 XX  
 PD 22-FEB-1996.  
 XX



```

PF 11-AUG-1995; 95MO-US10202.
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
XX
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eldens DM;
XX WPI: 1996-139702/14.
XX N-PSDB: AAT17489.
XX
XX
XX
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX PT gene - for diagnosis and therapy of human breast and ovarian cancer
XX PT gene - for diagnosing pre-disposition to these cancers
XX
XX
XX Claim 1; ; 218bp; English.
XX
XX
XX AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
XX CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
XX CC (see AAR81481 for wild type protein). These mutations can be used as
XX CC immunogens for antibody production. The mutant BRCA1 genes encoding
XX CC these sequences have at least 1 mutation or polymorphism in comparision
XX CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
XX CC gemiline alteration in the wild type BRCA1 gene, a predisposition for
XX CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
XX CC isolated from a tissue sample from a subject has a probe, corresponding
XX CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
XX CC allele-specific probe for a mutation of it), added to it. The conditions
XX CC allow for hybridisation of the probe to the mRNA, and any hybridisation
XX CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
XX CC sample is isolated, and a shift in electrophoretic mobility of single
XX CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
XX CC indicates a mutation. These methods of detection can also diagnose a
XX CC lesion neoplasia associated with the BRCA1 locus. The methods may be
XX CC used in gene therapy, protein replacement therapy and protein mimetics,
XX CC and may be used to screen for drugs in cancer therapy.
XX
XX
XX Sequence 1863 AA:
XX
XX
XX Query Match 99.9%; Score 9638; DB 17; Length 1863;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0.
XX
XX
XX 1 MDLSALRREEVONINAMOKLTLECPICELIKEPSTKCDHIFCFMCKLTLNOKKGSQ 60
XX |
XX 1 MDLSALRREEVONINAMOKLTLECPICELIKEPSTKCDHIFCFMCKLTLNOKKGSQ 60
XX
XX 61 CPLCKNDITRKSLOESTRFSSQVLEELKLTICAFOLDLLEYANSYNFAKKENSPENHLKD 120
XX |
XX 61 CPLCKNDITRKSLOESTRFSSQVLEELKLTICAFOLDLLEYANSYNFAKKENSPENHLKD 120
XX
XX 121 EVSITIOSMGYRNRAKRLLOESPENPSLOETSLSVOLSNLGTVRLTRKQRIOPKTSYVI 180
XX |
XX 121 EVSITIOSMGYRNRAKRLLOESPENPSLOETSLSVOLSNLGTVRLTRKQRIOPKTSYVI 180
XX
XX 181 ELGSDSSEDPYNRKATYCSVGOELLQITPOSTRDEISLDSAKKACSESDVYNTENHQ 240
XX |
XX 181 ELGSDSSEDPYNRKATYCSVGOELLQITPOSTRDEISLDSAKKACSESDVYNTENHQ 240
XX
XX 241 PSNNDLMTTEKRAAREHPEKTYOGSSVSNLHVEPCGTINHAASLOHENSLLITKDRMAYE 300
XX |
XX 241 PSNNDLMTTEKRAAREHPEKTYOGSSVSNLHVEPCGTINHAASLOHENSLLITKDRMAYE 300
XX
XX 301 KAEFCNKSQOPGLARSQHNRMWAGSKETCNDRRPTSTKRVKVLNDPLCEKRWNKOKLPC 360
XX |
XX 301 KAEFCNKSQOPGLARSQHNRMWAGSKETCNDRRPTSTKRVKVLNDPLCEKRWNKOKLPC 360

```

Dh	301	KAEPNKSQOGLASQNNRRAQSKETPCNDRTBETSEKKVLDMLDPJLCEBKEKMNKQKLP	360
Qy	361	SENPRDTEDEVPIUTLNSSIQKVENWFSKSDDELLGSDSDHDESESNARVADVLADLYEVD	420
Dh	361	SENPRDTEDEVPIUTLNSSIQKVENWFSKSDDELLGSDSDHDESESNARVADVLADLYEVD	420
Qy	421	EYSGSSEKIDLLASDPHALICKSERVHKSYESNIEDKIGKITYRKKAASLPNLSHVTEN	480
Dh	421	EYSGSSEKIDLLASDPHALICKSERVHKSYESNIEDKIGKITYRKKAASLPNLSHVTEN	480
Qy	481	LIIGFVFEPOIOIORPLTNKLRKRRTSGJHPEDFIKKDIAVOQKPEMNQSTNOTE	540
Dh	481	LIIGFVFEPOIOIORPLTNKLRKRRTSGJHPEDFIKKDIAVOQKPEMNQSTNOTE	540
Qy	541	ONGQVMNTNSNGHEKTKGDSIQNKKNNPIESLEKESAFYTKAPJISSISNMLEANI	600
Dh	541	ONGQVMNTNSNGHEKTKGDSIQNKKNNPIESLEKESAFYTKAPJISSISNMLEANI	600
Qy	601	HNKSAKKNNLRKRSSTHIALELVSRNLSPNCELTQIDSCSSDEIKKKKYNQPV	660
Dh	601	HNKSAKKNNLRKRSSTHIALELVSRNLSPNCELTQIDSCSSDEIKKKKYNQPV	660
Qy	661	RHSRLQIDMEKEPRTGAKSKNKPBEQJSKRHDOSTPELTKLTNMPGSEFTCSNSELKE	720
Dh	661	RHSRLQIDMEKEPRTGAKSKNKPBEQJSKRHDOSTPELTKLTNMPGSEFTCSNSELKE	720
Qy	721	FVNPSLPREKEEKELETYKVSNNMADPPDMLMSGERTVQTERSVSSSISLVPCTDYGTQ	780
Dh	721	FVNPSLPREKEEKELETYKVSNNMADPPDMLMSGERTVQTERSVSSSISLVPCTDYGTQ	780
Qy	781	ESISLLEVSTIGKAKTEPNKCVSOCAFENPKGLIHGCSKDNRDTEBGFKPYLGEVHNS	840
Dh	781	ESISLLEVSTIGKAKTEPNKCVSOCAFENPKGLIHGCSKDNRDTEBGFKPYLGEVHNS	840
Qy	841	RETSIEMESELDAQYLONTFKVSKROSEFALFJNSNGNMBEBCATPSAHSGSLKQSPVYT	900
Dh	841	RETSIEMESELDAQYLONTFKVSKROSEFALFJNSNGNMBEBCATPSAHSGSLKQSPVYT	900
Qy	901	FECEQKEKNQCKNESNIKPVQTVNTTAQFPVYVQGDKRPVDNAKGSIKGSRFCLSSQCRG	960
Dh	901	FECEQKEKNQCKNESNIKPVQTVNTTAQFPVYVQGDKRPVDNAKGSIKGSRFCLSSQCRG	960
Qy	961	NETGLITPNKGLLONPYRIPPLPIKSPVTKCKNLLENFEBHMSPEREMGENEIP	1020
Dh	961	NETGLITPNKGLLONPYRIPPLPIKSPVTKCKNLLENFEBHMSPEREMGENEIP	1020
Qy	1021	STVSTISNNIRVENVFKAASSNINEVSSNINEGSSNINEGSSDENIOAELGRNRGKL	1080
Dh	1021	STVSTISNNIRVENVFKAASSNINEVSSNINEGSSNINEGSSDENIOAELGRNRGKL	1080
Qy	1081	NAMLRGLQJLOPEVYKQSLPGSNCKHPETIKQOEYEVQVYVMTDESPYLISDNLEDPKMS	1140
Dh	1081	NAMLRGLQJLOPEVYKQSLPGSNCKHPETIKQOEYEVQVYVMTDESPYLISDNLEDPKMS	1140
Qy	1141	HASQVCSTPDLLDDGEIKEDTSAFENDIKESSAVFSKSVQKGLSRSPSPTHTHLAQ	1200
Dh	1141	HASQVCSTPDLLDDGEIKEDTSAFENDIKESSAVFSKSVQKGLSRSPSPTHTHLAQ	1200
Qy	1201	GYRGAKKLESSENLSEDELEPCFOHLLFGXVNNIPQOSTRHSYVATELSKNTENL	1260
Dh	1201	GYRGAKKLESSENLSEDELEPCFOHLLFGXVNNIPQOSTRHSYVATELSKNTENL	1260
Qy	1261	LSLKNLSLDCSNQVYLAKASQEHNLSEETKCSASLFSQCSSELEDILTANTNTPOPFLGS	1320
Dh	1261	LSLKNLSLDCSNQVYLAKASQEHNLSEETKCSASLFSQCSSELEDILTANTNTPOPFLGS	1320
Qy	1321	SKQMRHOSQGVGLSDKEIVSDDEBRCGTGLEENNQBQSDMSNLGEAASCESETSYSE	1380
Dh	1321	SKQMRHOSQGVGLSDKEIVSDDEBRCGTGLEENNQBQSDMSNLGEAASCESETSYSE	1380
Qy	1381	DCSGLSQSDILTTQOQRTQOHNLIKIQOEAELAVLEQHGSPSNYSPIITDSASLE	1440
Dh	1381	DCSGLSQSDILTTQOQRTQOHNLIKIQOEAELAVLEQHGSPSNYSPIITDSASLE	1440

QY	1441	DLRNPQOSTSEKAVLTSOKSSEYPIQONPEGLSADKFEVSADSTSKNKEPGYERSSPSK	1500
Db	1441	DLRNPQOSTSEKAVLTSOKSSEYPIQONPEGLSADKFEVSADSTSKNKEPGYERSSPSK	1500
QY	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYPRLDLEG	1560
Db	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYPRLDLEG	1560
QY	1561	TPYLESGLTFSDDPSPSEDRAPEARVGNIPSSTSALKVPOLKVAESAQSPAAAHHT	1620
Db	1561	TPYLESGLTFSDDPSPSEDRAPEARVGNIPSSTSALKVPOLKVAESAQSPAAAHHT	1620
QY	1621	DTAGYNAMEESYSREKPELTASTERYNKRMSVWSGLTPEEFMLVYKFAKHHITLTNLI	1680
Db	1621	DTAGYNAMEESYSREKPELTASTERYNKRMSVWSGLTPEEFMLVYKFAKHHITLTNLI	1680
QY	1681	TEETHVVMKTDAEFYVCERTLKIFLGIAGKVVSYFWVTOSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETHVVMKTDAEFYVCERTLKIFLGIAGKVVSYFWVTOSIKERKMLNEHDFEVRGDV	1740
QY	1741	VNGRHHOGPKRARESDRKFIFGLFETCCYGPFTNMPTDLEMMVOLCGASVVKELSSFTL	1800
Db	1741	VNGRHHOGPKRARESDRKFIFGLFETCCYGPFTNMPTDLEMMVOLCGASVVKELSSFTL	1800
QY	1801	GTGVHPVVOPDAMTEDNGFHAIGMCAPVVTREMLDVSVALYOCOEIDTYLIPQIPH	1860
Db	1801	GTGVHPVVOPDAMTEDNGFHAIGMCAPVVTREMLDVSVALYOCOEIDTYLIPQIPH	1860
QY	1861	SHY 1863	
Db	1861	SHY 1863	

Search completed: June 27, 2003, 10:22:32  
 Job time : 100 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 27, 2003, 10:20:56 ; Search time 45 Seconds  
(without alignments)  
4539.379 Million cell updates/sec

Title: US-09-734-672-4  
Perfect score: 9649  
Sequence: 1 MDLSALRVEYQVYINAMQK.....LYCCQELDTYLIQIPHSY 1863

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters: 424699

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCr\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/PCrUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9649	100.0	1863	US-09-734-672-4	Sequence 4, Appli
2	9649	100.0	1863	US-09-982-828-6	Sequence 6, Appli
3	9635	99.9	1863	US-09-734-672-2	Sequence 2, Appli
4	9635	99.9	1863	US-09-734-672-6	Sequence 6, Appli
5	9635	99.9	1863	US-09-982-828-2	Sequence 2, Appli
6	9635	99.9	1863	US-09-982-828-4	Sequence 4, Appli
7	9627	99.8	1863	US-10-022-819-2	Sequence 2, Appli
8	357	3.7	2344	US-09-815-242-12713	Sequence 12713, A
9	341	3.5	6281	US-09-815-242-12996	Sequence 12996, A
10	333	3.5	3899	US-10-171-311-4	Sequence 4, Appli
11	333	3.5	3917	US-10-171-311-8	Sequence 8, Appli
12	329	3.4	3907	US-10-171-311-2	Sequence 2, Appli
13	329	3.4	3925	US-10-171-311-6	Sequence 6, Appli
14	328.5	3.4	2368	US-09-815-242-5635	Sequence 5635, Ap
15	328.5	3.4	2368	US-09-815-242-12389	Sequence 12389, A
16	314	3.3	1596	US-09-902-432-4	Sequence 4, Appli
17	312	3.2	1400	US-09-764-176-7	Sequence 7, Appli
18	312	3.2	2478	US-09-815-242-5816	Sequence 5816, Ap
19	312	3.2	2478	US-09-815-242-12967	Sequence 12967, A

20	308	3.2	2665	US-09-864-761-34248	Sequence 34248, A
21	306	3.2	3158	US-09-815-242-12611	Sequence 12611, A
22	299.5	3.1	2843	US-09-987-482-1	Sequence 1, Appli
23	298.5	3.1	2843	US-08-681-219-32	Sequence 32, Appli
24	289	3.0	1979	US-10-205-823-419	Sequence 419, App
25	288	3.0	2476	US-09-824-574-7	Sequence 7, Appli
26	286.5	3.0	1781	US-09-961-403-13	Sequence 13, Appli
27	286.5	3.0	1781	US-09-738-877-3	Sequence 3, Appli
28	285.5	3.0	2025	US-09-815-242-5703	Sequence 5703, Ap
29	282	2.9	50	US-09-998-667-15	Sequence 15, Appli
30	278	2.9	3256	US-09-919-039-21	Sequence 21, Appli
31	278	2.9	3256	US-09-919-172-98	Sequence 98, Appli
32	272	2.8	2437	US-09-815-242-5834	Sequence 5834, Ap
33	268	2.8	5795	US-09-815-242-12610	Sequence 12610, A
34	267.5	2.8	1664	US-09-978-343-2	Sequence 2, Appli
35	267.5	2.8	1664	US-09-964-858-1	Sequence 1, Appli
36	264	2.7	1639	US-10-087-464-10	Sequence 10, Appli
37	262.5	2.7	1312	US-10-071-179-29	Sequence 29, Appli
38	261	2.7	1285	US-09-982-091A-2	Sequence 2, Appli
39	261	2.7	1938	US-10-171-311-164	Sequence 164, App
40	261	2.7	1945	US-09-927-597-2	Sequence 2, Appli
41	259	2.7	2383	US-10-082-830-260	Sequence 260, App
42	258.5	2.7	1972	US-10-171-311-162	Sequence 162, App
43	258.5	2.7	1979	US-09-927-597-4	Sequence 4, Appli
44	258	2.7	1786	US-09-742-096-3	Sequence 3, Appli
45	256	2.7	1616	US-09-820-843A-16	Sequence 16, Appli

## ALIGNMENTS

RESULT 1  
US-09-734-672-4  
Sequence 4, Application US/09734672  
Publication No. US20020183268A1

## GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin

TITLE OF INVENTION: Coding Sequences of the Human  
BRCAL Gene

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington

STATE: District of Columbia  
COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96

ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan

REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000

TELEFAX: 202-739-3001

```

: INFORMATION FOR SEQ ID NO: 4:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 1863 amino acids
:     TYPE: amino acid
:     STRANDEDNESS: No. US20020183268A1 Relevant
:     TOPOLOGY: No. US20020183268A1 Relevant
:     MOLECULE TYPE: protein
:     ORIGINAL SOURCE:
:       ORGANISM: Homo sapiens
:       STRAIN: BRCAl
:     POSITION IN GENOME:
:       CHROMOSOME/SEGMENT: 17
:       MAP POSITION: 17q21
:   SEQUENCE DESCRIPTION: SEQ ID NO: 4:
: US-09-734-672-4

Query Match      100.0%: Score 9649; DB 9: Length 1863;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MDLSALRVEQVQVIVINAMOKIIECPICLELKEPEVSTKCDHIFCKFCMKLKLNOCKGPSQ 60
DB      1 MDLSALRVEQVQVIVINAMOKIIECPICLELKEPEVSTKCDHIFCKFCMKLKLNOCKGPSQ 60
QY      61 CPLCKNDITKRSLOESTRFESQVLEELIKITICAFQOLDTGLEANSYNFAKKENNSPEHLKD 120
DB      61 CPLCKNDITKRSLOESTRFESQVLEELIKITICAFQOLDTGLEANSYNFAKKENNSPEHLKD 120
QY      121 EYSIIOSMGYRNRAKRLLOSEPENPSLOETSIVOLSNIGTVTLTKTKOIQOKSVYI 180
DB      121 EYSIIOSMGYRNRAKRLLOSEPENPSLOETSIVOLSNIGTVTLTKTKOIQOKSVYI 180
QY      181 ELGSDSEDTYVKATYCSYDQDELLQITQGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
DB      181 ELGSDSEDTYVKATYCSYDQDELLQITQGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
QY      241 PENNDLNTTEKRAAEHPEKYQSSVSNLHVEPCGNTNTHASSLOHENSLLTKDRMNV 300
DB      241 PENNDLNTTEKRAAEHPEKYQSSVSNLHVEPCGNTNTHASSLOHENSLLTKDRMNV 300
QY      301 KAEFCCKSKOPGLARSOHNRMAKSKETCDNDRTPTSEKKVYDLNADPLCEKKNKOLPC 360
DB      301 KAEFCCKSKOPGLARSOHNRMAKSKETCDNDRTPTSEKKVYDLNADPLCEKKNKOLPC 360
QY      361 SENPRDEDPVMTLNSSIQKVNEFMSRSDDELLGSDSHDGESESNAKYADVLDVLENDV 420
DB      361 SENPRDEDPVMTLNSSIQKVNEFMSRSDDELLGSDSHDGESESNAKYADVLDVLENDV 420
QY      421 EYSGSSEKIDLLASDPHEALICKSEYVHSHKSVESNIEDKIFGKTYRKKASLPMLSHVTEN 480
DB      421 EYSGSSEKIDLLASDPHEALICKSEYVHSHKSVESNIEDKIFGKTYRKKASLPMLSHVTEN 480
QY      481 LITGAFTVEPQIIQERPLTNKLKRRKRPISGLHPEDFIKKADLAVOKTPETMIQGNQTE 540
DB      481 LITGAFTVEPQIIQERPLTNKLKRRKRPISGLHPEDFIKKADLAVOKTPETMIQGNQTE 540
QY      541 ONGQVYMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSISIMMELELN 600
DB      541 ONGQVYMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSISIMMELELN 600
QY      601 HNSKAPKRNRLRKSSTRIHALELVYRNLSPPNCTEIQIDSCSSSEELKAKKYYQMVEY 660
DB      601 HNSKAPKRNRLRKSSTRIHALELVYRNLSPPNCTEIQIDSCSSSEELKAKKYYQMVEY 660
QY      661 RHRSMIQLMEKKEPATGAKKSNKPNQTSKRDHSDPEFPELKLITNAGSTKCSNTSELKE 720
DB      661 RHRSMIQLMEKKEPATGAKKSNKPNQTSKRDHSDPEFPELKLITNAGSTKCSNTSELKE 720
QY      721 FVNPSLPREEKEKELETYVSNNAEDPKDMLSGEYVLOTERTSESSSISLVPGTGYGTO 780
DB      721 FVNPSLPREEKEKELETYVSNNAEDPKDMLSGEYVLOTERTSESSSISLVPGTGYGTO 780
QY      781 ESIISLEVSTLGAKTEPNKCVSQAFAFENPKGLIHGCSKDNKNDTEGFKYPLGHEVNH 840
DB      781 ESIISLEVSTLGAKTEPNKCVSQAFAFENPKGLIHGCSKDNKNDTEGFKYPLGHEVNH 840
```

```

DB      841 RETSIMESELDIAOVLONTFKYKRSQFALSNSGMAEEBCATFSANSGSLKJKSPRYT 900
QY      841 RETSIMESELDIAOVLONTFKYKRSQFALSNSGMAEEBCATFSANSGSLKJKSPRYT 900
DB      901 FECEOKEENQKNESNIRKPVQVNTIAGFPVVGOKDKPYDNKACSIKSGSRECLSSQPRG 960
QY      901 FECEOKEENQKNESNIRKPVQVNTIAGFPVVGOKDKPYDNKACSIKSGSRECLSSQPRG 960
DB      961 NETGLITPNKHGGLQNPYRIPLPLPIKSFVTKCKKNILEENFEESHMSPEREMGNENIP 1020
QY      961 NETGLITPNKHGGLQNPYRIPLPLPIKSFVTKCKKNILEENFEESHMSPEREMGNENIP 1020
DB      1021 STVSTISRNINIRNVKREASSNININVGSTNEVSSINELGSSDENITQAEIGRRNGPKL 1080
QY      1021 STVSTISRNINIRNVKREASSNININVGSTNEVSSINELGSSDENITQAEIGRRNGPKL 1080
DB      1081 NAMLRLGLVLOPEYVKQSLPGSNCKHPEIKKQYEVEVQTVNTDFSPYLLSDNLEQPMGSS 1140
QY      1081 NAMLRLGLVLOPEYVKQSLPGSNCKHPEIKKQYEVEVQTVNTDFSPYLLSDNLEQPMGSS 1140
DB      1141 HASQVCESTPDDLDDGEIKEDTSPAENDIKESSAVFSKSYQKGLSRSPPFTHTHLAQ 1200
QY      1141 HASQVCESTPDDLDDGEIKEDTSPAENDIKESSAVFSKSYQKGLSRSPPFTHTHLAQ 1200
DB      1201 GYRGAKKLLESSPENLSEDEELPCQHLLPCKVNNIPSOGRHSTVATECLSKNTEENL 1260
QY      1201 GYRGAKKLLESSPENLSEDEELPCQHLLPCKVNNIPSOGRHSTVATECLSKNTEENL 1260
DB      1261 LSLKNSLNDCSNOVILIAKASQEHLSSEETKCSASLFFSSQCSLELDTANTMTOPFLIGS 1320
QY      1261 LSLKNSLNDCSNOVILIAKASQEHLSSEETKCSASLFFSSQCSLELDTANTMTOPFLIGS 1320
DB      1321 SKOMRHQSSQGVGLSDKELVSDDEERGTLGEENNOEBSMDNUGEASGCESETSVSE 1380
QY      1321 SKOMRHQSSQGVGLSDKELVSDDEERGTLGEENNOEBSMDNUGEASGCESETSVSE 1380
DB      1381 DCSGLSSQSDILITTOQRDMQHNLIKLOEAMELEAVLEOHGSOPSNSYPSIISQSALE 1440
QY      1381 DCSGLSSQSDILITTOQRDMQHNLIKLOEAMELEAVLEOHGSOPSNSYPSIISQSALE 1440
DB      1441 DLRNPEQSTSEKAVILTSQKSSSEYPISQNEGLSADKFEVSADSSYTKKKEGVERSPPSK 1500
QY      1441 DLRNPEQSTSEKAVILTSQKSSSEYPISQNEGLSADKFEVSADSSYTKKKEGVERSPPSK 1500
DB      1501 CFSLDRLRWYHSCSGSLQNRNYPQOEELIKVYDVEEQULEESGPHDLTETSYLPRODLEG 1560
QY      1501 CFSLDRLRWYHSCSGSLQNRNYPQOEELIKVYDVEEQULEESGPHDLTETSYLPRODLEG 1560
DB      1561 TPYLESGISLFDSDPSPSEDRAPESARVGINPSTSLAKYPOLKVAESQSPAAATTT 1620
QY      1561 TPYLESGISLFDSDPSPSEDRAPESARVGINPSTSLAKYPOLKVAESQSPAAATTT 1620
DB      1621 DTAGYNAMEESVREKPELTASTERYNKRMSWVSGLTPPEEMLYKFKARKHHTLTNLI 1680
QY      1621 DTAGYNAMEESVREKPELTASTERYNKRMSWVSGLTPPEEMLYKFKARKHHTLTNLI 1680
DB      1681 TEETHYVAKTAAEYVCERTLKYFLIGIAGKAVVSYFWVTOSIKERKMLNHDFEVNGDV 1740
QY      1681 TEETHYVAKTAAEYVCERTLKYFLIGIAGKAVVSYFWVTOSIKERKMLNHDFEVNGDV 1740
DB      1741 VGNRNHQGPKRARESDRKIFRGELCCGPFPTNPTQOLEMMVOLCGASVYKELSSFTL 1800
QY      1741 VGNRNHQGPKRARESDRKIFRGELCCGPFPTNPTQOLEMMVOLCGASVYKELSSFTL 1800
DB      1801 GTGVHPPIVVQDPANTEDNGFHAIGOMCEAPVYTRFWVLDVVALYQCELDITYLIPQIPH 1860
QY      1801 GTGVHPPIVVQDPANTEDNGFHAIGOMCEAPVYTRFWVLDVVALYQCELDITYLIPQIPH 1860
DB      1861 SHY 1863
QY      1861 SHY 1863
```

Db 1861 SHY 1863

## RESULT 2

US-09-982-828-6

Sequence 6, Application US/09982828  
Publication No. US20030022184A1

## GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.  
Allen, Antonette C.  
Alvares, Christopher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Thurber, Denise  
Zeng, BinTITLE OF INVENTION: Coding Sequences of the Human  
BRCA1 Gene

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Avenue N. W.  
CITY: Washington  
STATE: DC

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/982,828

FILING DATE: 22-Oct-2001

CLASSIFICATION: &lt;unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/074,453

FILING DATE: 1998-05-06

APPLICATION NUMBER: US 08/798,691

FILING DATE: 1997-02-12

APPLICATION NUMBER: US 08/598,591

FILING DATE: 1996-02-12

ATTORNEY/AGENT INFORMATION:

NAME: Michael S. Tuscan

REGISTRATION NUMBER: 43,210

REFERENCE/DOCKET NUMBER: 44921-5053-01-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-739-3000

TELEFAX: 202-739-3001

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1863 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

STRAIN: BRCA1 (om13)

POSITION IN GENOME:

CHROMOSOME/SEGMENT: 17

MAP POSITION: 17q21

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-982-828-6

Query Match 100.0%; Score 9649; DB 9; Length 1863;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSALVEEVQVNVNAMOKILCEPICLELIKPEVSTKCDHIFCKFCMLKLLNOKKPSQ 60  
Db 1 MDLSALVEEVQVNVNAMOKILCEPICLELIKPEVSTKCDHIFCKFCMLKLLNOKKPSQ 60  
QY 61 CPLCKNDITKRSLOESTRFSQVLEELIKICAFOLDTGLEAVANSYNFAKKENNSPEHLKD 120

Db 61 CPLCKNDITKRSLOESTRFSQVLEELIKICAFOLDTGLEAVANSYNFAKKENNSPEHLKD 120

QY 121 EVSTIOSMGYRNRRKRLJOSEPEPNSLOETSLVSQVSLNCTVRLTRKORLOPOKTSVYI 180

Db 121 EVSTIOSMGYRNRRKRLJOSEPEPNSLOETSLVSQVSLNCTVRLTRKORLOPOKTSVYI 180

QY 181 ELGSDSSEEDVNVKATYCVSDQDELQITPGTDEISLDSAKRAACEFSETDVTNTEHHQ 240

Db 181 ELGSDSSEEDVNVKATYCVSDQDELQITPGTDEISLDSAKRAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVEPCGTNTAASLOHENSLLLTDRMNV 300

Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVEPCGTNTAASLOHENSLLLTDRMNV 300

QY 301 KAEFCNKSQPGRLARSHNNWAGSKETCNDRPTSTKRYDLNADPLCEKKNKOKLPC 360

Db 301 KAEFCNKSQPGRLARSHNNWAGSKETCNDRPTSTKRYDLNADPLCEKKNKOKLPC 360

QY 361 SENPRDTEDPWITLNSIQKVNEMFGRSDELLGSDSHGSEESNAKVADVLVNEVD 420

Db 361 SENPRDTEDPWITLNSIQKVNEMFGRSDELLGSDSHGSEESNAKVADVLVNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSBNIEDKIFGKTYRRKASLPNLSHTEN 480

Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSBNIEDKIFGKTYRRKASLPNLSHTEN 480

QY 481 LIIGAFTEPQIIIOERLTNKLKRRRPTSGIHPEDEIKRADLAVOTPEMINQNOTE 540

Db 481 LIIGAFTEPQIIIOERLTNKLKRRRPTSGIHPEDEIKRADLAVOTPEMINQNOTE 540

QY 541 QNGOVNMTNSGHENKTKGDSIQNEKNPNIESLEKESAKRTAEPISISSINMTELENI 600

Db 541 QNGOVNMTNSGHENKTKGDSIQNEKNPNIESLEKESAKRTAEPISISSINMTELENI 600

QY 601 HNSKAPKKNRLRRKSSSTRHIALELVVSRNLSPPNCTEIQIDSCSSSEETIKKKYQMPV 660

Db 601 HNSKAPKKNRLRRKSSSTRHIALELVVSRNLSPPNCTEIQIDSCSSSEETIKKKYQMPV 660

QY 661 RHSRNQLMGKRPATGAKKSNPNEDTSKRHSDTPELKLTNAPSPFKCSTSLKE 720

Db 661 RHSRNQLMGKRPATGAKKSNPNEDTSKRHSDTPELKLTNAPSPFKCSTSLKE 720

QY 721 FVNPSPLEKEKEKLETVKVSNNAEDEPKDMLGGERVLOTREVSSESSISLVPGTDGTQ 780

Db 721 FVNPSPLEKEKEKLETVKVSNNAEDEPKDMLGGERVLOTREVSSESSISLVPGTDGTQ 780

QY 781 ESISLLEVSTLGAKTPEPNKCVSQAFAFENPKGLIHGCSKDNBNDEGFRYPLGHEVNH 840

Db 781 ESISLLEVSTLGAKTPEPNKCVSQAFAFENPKGLIHGCSKDNBNDEGFRYPLGHEVNH 840

QY 841 RETSIEMESELDQVLYONTFKVSKROSFALEFNPNGAEBCATFSAHSGSLKOSKVT 900

Db 841 RETSIEMESELDQVLYONTFKVSKROSFALEFNPNGAEBCATFSAHSGSLKOSKVT 900

QY 901 FECEOKKEENGKKNESNKPQVNTNITAGFPVVOCKDPVYNACSLIKGSRFLCSSQFRG 960

Db 901 FECEOKKEENGKKNESNKPQVNTNITAGFPVVOCKDPVYNACSLIKGSRFLCSSQFRG 960

QY 961 NETGLITPNKHGLLQNPYRIPLLEPIKSFVTKCKKNLLEENFEHSMSPEREMGENIP 1020

Db 961 NETGLITPNKHGLLQNPYRIPLLEPIKSFVTKCKKNLLEENFEHSMSPEREMGENIP 1020

QY 1021 STYSTISRNIRRENVREKASSNINEVGSSSTNEVGSSINTEGSDDEITQALGRNGPKL 1080

Db 1021 STYSTISRNIRRENVREKASSNINEVGSSSTNEVGSSINTEGSDDEITQALGRNGPKL 1080

QY 1081 NAMLRGLVLOPEYKQSLPSCNCKHPEIKKQEEVVOVNTPEFSPYLLSDNLEQMGSS 1140

Db 1081 NAMLRGLVLOPEYKQSLPSCNCKHPEIKKQEEVVOVNTPEFSPYLLSDNLEQMGSS 1140

QY 1141 HASOVSETPDDLLDQGETIKEDTSAENDIKESSAAVFSKVOGELSRSPPTTHLQO 1200

Db 1141 HASOVSETPDDLLDQGETIKEDTSAENDIKESSAAVFSKVOGELSRSPPTTHLQO 1200





ORGANISM: Homo sapiens  
 STRAIN: BRCA1  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 17  
 MAP POSITION: 17q21  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-734-672-6

Query Match 99.9% Score 9635; DB 9; Length 1863;  
 Best Local Similarity 99.8% Pred. No. 0;  
 Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 MDLSALRYEEVONVINAQKILECPICLLEIKPVSCTDHIKCFKCMKLLNÖKKGPQ 60
DB 1 MDLSALRYEEVONVINAQKILECPICLLEIKPVSCTDHIKCFKCMKLLNÖKKGPQ 60
OY 61 CPLCNDITKRSLOSTRFSOLVEELKTIICAFOLDTGLEYANSYNPAKKENSPHLD 120
DB 61 CPLCNDITKRSLOSTRFSOLVEELKTIICAFOLDTGLEYANSYNPAKKENSPHLD 120
OY 121 EYSIIQSMGYRRARLLÖSEPEPNSLOETSLVSÖLNGTVRTLRKQRIÖPQKTSVYI 180
DB 121 EYSIIQSMGYRRARLLÖSEPEPNSLOETSLVSÖLNGTVRTLRKQRIÖPQKTSVYI 180
OY 181 ELGSSSEEDTVNKATYCSVGOELQITPQGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
DB 181 ELGSSSEEDTVNKATYCSVGOELQITPQGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
OY 241 PSNNDLNTEKRAAERHPEKYGSSSVSNLHVEPCGNTNTHASSLÖHNSSLLTKDMANYE 300
DB 241 PSNNDLNTEKRAAERHPEKYGSSSVSNLHVEPCGNTNTHASSLÖHNSSLLTKDMANYE 300
OY 301 KAEFCNKSÖPGLARSÖHNRMAGSKETCNDRTPTSEKRVLDNADPLCERKENKÖKLCPC 360
DB 301 KAEFCNKSÖPGLARSÖHNRMAGSKETCNDRTPTSEKRVLDNADPLCERKENKÖKLCPC 360
OY 361 SENPDTEDEVPWITLNSIQKYNEMFSRDELLGSDSHDGESESAKAVADVDVNEVD 420
DB 361 SENPDTEDEVPWITLNSIQKYNEMFSRDELLGSDSHDGESESAKAVADVDVNEVD 420
OY 421 EYSSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYKKASLPVLVHTEN 480
DB 421 EYSSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYKKASLPVLVHTEN 480
OY 481 LIIGAFTVEPÖIÖERPLTNKLRKRRTSGLHPEDFIKKADLAVÖKTEPMINÖGTNOTE 540
DB 481 LIIGAFTVEPÖIÖERPLTNKLRKRRTSGLHPEDFIKKADLAVÖKTEPMINÖGTNOTE 540
OY 541 ÖNGÖVMNTNSGHEKKTGDSIQÖNEKNPPIESLEKESAFKTAEPISISSIMLELENI 600
DB 541 ÖNGÖVMNTNSGHEKKTGDSIQÖNEKNPPIESLEKESAFKTAEPISISSIMLELENI 600
OY 601 HNSKAPKNRLRKRSTRIHLELVSNLSPPNCTELÖIDSCSSEETKKKKYNÖMPV 660
DB 601 HNSKAPKNRLRKRSTRIHLELVSNLSPPNCTELÖIDSCSSEETKKKKYNÖMPV 660
OY 661 RHSRNLÖMEGEPATGAKKSNKPÖQTSKRHSDTFPELKLTNAGSESTKCSNTSELKE 720
DB 661 RHSRNLÖMEGEPATGAKKSNKPÖQTSKRHSDTFPELKLTNAGSESTKCSNTSELKE 720
OY 721 FVNPSLPREEKEKLEYKVSNNADPKDMLSGERVLOTERSVSSISLVPGTDYGRQ 780
DB 721 FVNPSLPREEKEKLEYKVSNNADPKDMLSGERVLOTERSVSSISLVPGTDYGRQ 780
OY 781 ESIISLLEVSTLGKATEPKNKVSÖCAAFENPKGLIHGCKDNRNDEGKRYPLGHEVNS 840
DB 781 ESIISLLEVSTLGKATEPKNKVSÖCAAFENPKGLIHGCKDNRNDEGKRYPLGHEVNS 840
OY 841 RETSIEMESELDAÖYLONTFEVSKROSÖFALFSNPGNAEECATSAHSGSLKÖSPKYT 900
DB 841 RETSIEMESELDAÖYLONTFEVSKROSÖFALFSNPGNAEECATSAHSGSLKÖSPKYT 900
OY 901 FECEÖKEENÖGNESNIKPVÖTÖVNTAGFPVVGÖKDPVNAKCSIKGSRFCSSÖFRG 960
  
```

```

DB 901 FECEÖKEENÖGNESNIKPVÖTÖVNTAGFPVVGÖKDPVNAKCSIKGSRFCSSÖFRG 960
OY 961 NETGLITPKKHGLOVPYRIPLPPIKPSVYKCKCKKNLLEEFHSHSPREKMNENIP 1020
DB 961 NETGLITPKKHGLOVPYRIPLPPIKPSVYKCKCKKNLLEEFHSHSPREKMNENIP 1020
OY 1021 STVSTISRNINIRENVFKEASSNINEVGSSTNEVSSINEIGSSDENIÖAELGRNRGPKL 1080
DB 1021 STVSTISRNINIRENVFKEASSNINEVGSSTNEVSSINEIGSSDENIÖAELGRNRGPKL 1080
OY 1081 NAMLRLGVLOPEYKÖSLPGSNCKHPEIKÖEYEEVOÖVNTDESPLYISDNLEÖPMGSS 1140
DB 1081 NAMLRLGVLOPEYKÖSLPGSNCKHPEIKÖEYEEVOÖVNTDESPLYISDNLEÖPMGSS 1140
OY 1141 HASÖVSETPDDLDGCEIKEDTSAENDIKRSSAVFSKSVÖKGELSPSPFTHTHQAQ 1200
DB 1141 HASÖVSETPDDLDGCEIKEDTSAENDIKRSSAVFSKSVÖKGELSPSPFTHTHQAQ 1200
OY 1201 GYRGAKKLESSEENLSEDEELPCFÖHLLFGKVNINISÖSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSEDEELPCFÖHLLFGKVNINISÖSTRHSTVATECLSKNTEENL 1260
OY 1261 LSLKNSLNDCSNOVYILAKASÖEHHLSEETKCSASLFSÖCÖSELBDLTANTÖPFLIGS 1320
DB 1261 LSLKNSLNDCSNOVYILAKASÖEHHLSEETKCSASLFSÖCÖSELBDLTANTÖPFLIGS 1320
OY 1321 SKÖMRHÖSESÖGVLSDKELVSDDEERGTLGLENÖBÖSMDSNUGELASÖCESSTVSSE 1380
DB 1321 SKÖMRHÖSESÖGVLSDKELVSDDEERGTLGLENÖBÖSMDSNUGELASÖCESSTVSSE 1380
OY 1381 DCSGSSÖSDILITTOÖQRTMÖHNLIKLÖEMAELEAVLEÖHGÖSPNSYPSIISSÖSALE 1440
DB 1381 DCSGSSÖSDILITTOÖQRTMÖHNLIKLÖEMAELEAVLEÖHGÖSPNSYPSIISSÖSALE 1440
OY 1441 DLRNPEÖQTSSEKAVILTSÖKSSEYPTISÖNPEGLSARKFEVSADSSTSKKKEGVENSFSK 1500
DB 1441 DLRNPEÖQTSSEKAVILTSÖKSSEYPTISÖNPEGLSARKFEVSADSSTSKKKEGVENSFSK 1500
OY 1501 CPSLDRMYMHSCSGSLÖNRNYPÖSÖBELIKVYDVEÖQLEESGPHDLTETSYLPPRODLEG 1560
DB 1501 CPSLDRMYMHSCSGSLÖNRNYPÖSÖBELIKVYDVEÖQLEESGPHDLTETSYLPPRODLEG 1560
OY 1561 TPYLESÖISLFSDDPESDPSEDRAPEASRVGNIPSTSAKLVÖLKVAESASÖPAAHTT 1620
DB 1561 TPYLESÖISLFSDDPESDPSEDRAPEASRVGNIPSTSAKLVÖLKVAESASÖPAAHTT 1620
OY 1621 DTAGYNAMESVSREKPELTLASTERYNKRMSVWSGLTPPEERMLYKFARRHHITLNL 1680
DB 1621 DTAGYNAMESVSREKPELTLASTERYNKRMSVWSGLTPPEERMLYKFARRHHITLNL 1680
OY 1681 TEETTHVVKTDAAEFYCERETLKYFLGIGAGKVVVSYPWVÖSIRKRMKLNDEHFEVRGDV 1740
DB 1681 TEETTHVVKTDAAEFYCERETLKYFLGIGAGKVVVSYPWVÖSIRKRMKLNDEHFEVRGDV 1740
OY 1741 VNGRNHÖGPKRARESÖDRKIFRGLEICCYGPTNPTDÖLEMMVÖLGCASVVKELSSFTL 1800
DB 1741 VNGRNHÖGPKRARESÖDRKIFRGLEICCYGPTNPTDÖLEMMVÖLGCASVVKELSSFTL 1800
OY 1801 GTGVNPIVVÖDANTEDNGFHAIGÖNCEAPVYTHENVLDSVALYÖCÖELDTYLIPOIPH 1860
DB 1801 GTGVNPIVVÖDANTEDNGFHAIGÖNCEAPVYTHENVLDSVALYÖCÖELDTYLIPOIPH 1860
OY 1861 SHY 1863
DB 1861 SHY 1863
  
```

RESULT 5  
 US-09-982-828-2  
 ; Sequence 2. Application us/09982828  
 ; Publication No. US20030022184A1  
 ; GENERAL INFORMATION:



APPLICANT: Murphy, Patricia D.  
 Allen, Antonette C.  
 Alvares, Christopher P.  
 Critz, Brenda S.  
 Olson, Sheri J.  
 Thurber, Denise  
 Zeng, Bin  
 TITLE OF INVENTION: Coding Sequences of the Human  
 BRCA1 Gene  
 NUMBER OF SEQUENCES: 72  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morgan Lewis & Bockius LLP  
 STREET: 1111 Pennsylvania Avenue N. W.  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/982,828  
 FILING DATE: 22-Oct-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/074,453  
 FILING DATE: 1998-05-06  
 APPLICATION NUMBER: US 08/798,691  
 FILING DATE: 1997-02-12  
 APPLICATION NUMBER: US 08/598,591  
 FILING DATE: 1996-02-12  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michael S. Tuscan  
 REGISTRATION NUMBER: 43,210  
 REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-739-3000  
 TELEFAX: 202-739-3001  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1863 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 STRAIN: BRCA1 (om11)  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 17  
 MAP POSITION: 17q21  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-982-828-2  
 Query Match 99.9%; Score 9635; DB 9; Length 1863;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 181 ELGSDSSEDTYVNAKATYCSVGDELLQITPOGTRDEISLDSAKKAACESEDTVTNTEHHQ 240  
 Qy 241 PSNNDLNTTEKRAAEHPEKRYOGSSVSNLJHEPGTNTASSLQHNESILLTDORNAVE 300  
 Db 241 PSNNDLNTTEKRAAEHPEKRYOGSSVSNLJHEPGTNTASSLQHNESILLTDORNAVE 300  
 Qy 301 KAEFCNKSQPGRLARSOHNHWAGSKETCNDRRTSTTEKRVLDNADPLCEKREWNKOKLPC 360  
 Db 301 KAEFCNKSQPGRLARSOHNHWAGSKETCNDRRTSTTEKRVLDNADPLCEKREWNKOKLPC 360  
 Qy 361 SENPRDTEVPWITLNSIQKVNEMFSRSDLLSDSDSHGDESSNAKAVADYDLNEVD 420  
 Db 361 SENPRDTEVPWITLNSIQKVNEMFSRSDLLSDSDSHGDESSNAKAVADYDLNEVD 420  
 Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVEENIEDKIKGTYRRKASLPNLSHTYEN 480  
 Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVEENIEDKIKGTYRRKASLPNLSHTYEN 480  
 Qy 481 LIIGAFYTEPQIIQERPLTNKLRKRRPTSGLHPEDFIKRADLAVOKTPEMINOGTNOTE 540  
 Db 481 LIIGAFYTEPQIIQERPLTNKLRKRRPTSGLHPEDFIKRADLAVOKTPEMINOGTNOTE 540  
 Qy 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLEKESAFYTKAEPITSSISSNMELEINI 600  
 Db 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLEKESAFYTKAEPITSSISSNMELEINI 600  
 Qy 601 HNSKAPKKNRRLRRKSSSTRHIALELVYSRNLSPPNCTELOTDCSSSSSEIKKKRYNQMPV 660  
 Db 601 HNSKAPKKNRRLRRKSSSTRHIALELVYSRNLSPPNCTELOTDCSSSSSEIKKKRYNQMPV 660  
 Qy 661 RHSRNLQIMEGKEPATGAKKSNKPNEDTSKRHSDTFPELKLITNAPGSGFTCSNTSEKE 720  
 Db 661 RHSRNLQIMEGKEPATGAKKSNKPNEDTSKRHSDTFPELKLITNAPGSGFTCSNTSEKE 720  
 Qy 721 FVNPDLPREKEKLETVKVSNNAEDEPKDMLSGERVLOTFRSVSSSISLVPGTDYGTQ 780  
 Db 721 FVNPDLPREKEKLETVKVSNNAEDEPKDMLSGERVLOTFRSVSSSISLVPGTDYGTQ 780  
 Qy 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENKGLIHGCSKNRNDTEGFKPLDHEVHVS 840  
 Db 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENKGLIHGCSKNRNDTEGFKPLDHEVHVS 840  
 Qy 841 RETSIEMESELDQOYLQNTFKVSKROSFALFSPNAGNAEECAFTSAHSGSLKROSPKVT 900  
 Db 841 RETSIEMESELDQOYLQNTFKVSKROSFALFSPNAGNAEECAFTSAHSGSLKROSPKVT 900  
 Qy 901 FECBQKEENQGNESNIRPVQTVNITAGFPVVGKDKRPVDNAKCSIKGSRFCLSSQFRG 960  
 Db 901 FECBQKEENQGNESNIRPVQTVNITAGFPVVGKDKRPVDNAKCSIKGSRFCLSSQFRG 960  
 Qy 961 NETGLITPNKHGLLQNTYRIRPLEPFIKSPYTKKKNULLENFEHSHSPREKNGENIP 1020  
 Db 961 NETGLITPNKHGLLQNTYRIRPLEPFIKSPYTKKKNULLENFEHSHSPREKNGENIP 1020  
 Qy 1021 STVSTISRNRIRENVFKEASSNINEGSSSTNEVGSSINIEGSSDENIOALGGRNRPKL 1080  
 Db 1021 STVSTISRNRIRENVFKEASSNINEGSSSTNEVGSSINIEGSSDENIOALGGRNRPKL 1080  
 Qy 1081 NAMRLGVLQPEVYKQSLPGSNCKHPEIKKQOEVEVQVNTDPSYLLISDNLEQPMGSS 1140  
 Db 1081 NAMRLGVLQPEVYKQSLPGSNCKHPEIKKQOEVEVQVNTDPSYLLISDNLEQPMGSS 1140  
 Qy 1141 HASOVCSSETDDLLDDEIKEDTSAFANDIKESAVSKSVQKELRSRSPFHTHLAQ 1200  
 Db 1141 HASOVCSSETDDLLDDEIKEDTSAFANDIKESAVSKSVQKELRSRSPFHTHLAQ 1200  
 Qy 1201 GYRGAKKLESSSEENLSSSEDEELPCFOHLLFGKANNIPPSOSTRSTYATFCLSKNTEENL 1260  
 Db 1201 GYRGAKKLESSSEENLSSSEDEELPCFOHLLFGKANNIPPSOSTRSTYATFCLSKNTEENL 1260  
 Qy 1261 LSLKNSINDCSNOYITAKASQEHHLSEETKCSASLFFSSQCSLEEDLTANTNTODPFLIGS 1320  
 Db 1261 LSLKNSINDCSNOYITAKASQEHHLSEETKCSASLFFSSQCSLEEDLTANTNTODPFLIGS 1320

Db 1261 LSLKNSLNDSCNOVILAKASQEHHLSEETKCSALFSSQCELEDLTANTNTODPELLIGS 1320  
 QY 1321 SKOMRHOSSESQVGLSDKELVSDDERGTGLLENNNOEODSMDSNLGEAASCESETSYSE 1380  
 Db 1321 SKOMRHOSSESQVGLSDKELVSDDERGTGLLENNNOEODSMDSNLGEAASCESETSYSE 1380  
 QY 1381 DCSGSSQSODILITTOORDTMOHNLKLOQMALEVLLEOHGQSPNSYPSIISDSALE 1440  
 Db 1381 DCSGSSQSODILITTOORDTMOHNLKLOQMALEVLLEOHGQSPNSYPSIISDSALE 1440  
 QY 1441 DLRNPQSTSEKAVLTSOKSSEYPISONPEGLSADAFVSAADSTSKNKPGRVRSPPSK 1500  
 Db 1441 DLRNPQSTSEKAVLTSOKSSEYPISONPEGLSADAFVSAADSTSKNKPGRVRSPPSK 1500  
 QY 1501 CPSLDDRWYMSQSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYPRDLEG 1560  
 Db 1501 CPSLDDRWYMSQSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYPRDLEG 1560  
 QY 1561 TPYLESGISLFSDDPESDSEDRAPESARVGNIPSSSTALKVPOLKVAESAQSPAAHTT 1620  
 Db 1561 TPYLESGISLFSDDPESDSEDRAPESARVGNIPSSSTALKVPOLKVAESAQSPAAHTT 1620  
 QY 1621 DTAGYNAMESVSREKPELTASTERYNKRMSVYSGLTPEEFMLVYKFAKHHTITLNL 1680  
 Db 1621 DTAGYNAMESVSREKPELTASTERYNKRMSVYSGLTPEEFMLVYKFAKHHTITLNL 1680  
 QY 1681 TEETHHVAKTDAEFYCERLTFLGAGKVVVSYFWYQSIKERKMLNEHDFEVRGVDY 1740  
 Db 1681 TEETHHVAKTDAEFYCERLTFLGAGKVVVSYFWYQSIKERKMLNEHDFEVRGVDY 1740  
 QY 1741 VNGRNHOGKRRARESDRKIFRGLEICCGPTNMPPTDLEWVYOLGASVVKELSFLL 1800  
 Db 1741 VNGRNHOGKRRARESDRKIFRGLEICCGPTNMPPTDLEWVYOLGASVVKELSFLL 1800  
 QY 1801 GTGVPIYVVDADWEDNGFHAIGOMCEAPVYTRFVLDVSVALYOCQELDTLLIPIPH 1860  
 Db 1801 GTGVPIYVVDADWEDNGFHAIGOMCEAPVYTRFVLDVSVALYOCQELDTLLIPIPH 1860  
 QY 1861 SHY 1863  
 Db 1861 SHY 1863

CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/074,453  
 FILING DATE: 1998-05-06  
 APPLICATION NUMBER: US 08/798,691  
 FILING DATE: 1997-02-12  
 APPLICATION NUMBER: US 08/598,591  
 FILING DATE: 1996-02-12  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michael S. Tuscan  
 REGISTRATION NUMBER: 43,210  
 REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-739-3000  
 TELEFAX: 202-739-3001  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1863 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 STRAIN: BRCA1 (om12)  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 17  
 MAP POSITION: 17q21  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
 US-09-982-828-4  
 Query Match 99.9%; Score 9635; DB 9; Length 1863;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MDLSARVEYQVNIAMQKILCEPICLELIEPEVSTKCDHIFCKCMKLLNKKGPSQ 60  
 Db 1 MDLSARVEYQVNIAMQKILCEPICLELIEPEVSTKCDHIFCKCMKLLNKKGPSQ 60  
 QY 61 CPICKNDITKRSLOESTRFSQLEVEELKILICAFOLDTGEVANSYVFAKKENNSEHLD 120  
 Db 61 CPICKNDITKRSLOESTRFSQLEVEELKILICAFOLDTGEVANSYVFAKKENNSEHLD 120  
 QY 121 EVSIIOGMYRNRAKRLQSEPNPSLOETSLSVLSNLGTVRTLRTKQRIQPKTSYVI 180  
 Db 121 EVSIIOGMYRNRAKRLQSEPNPSLOETSLSVLSNLGTVRTLRTKQRIQPKTSYVI 180  
 QY 181 ELGSDSEEDTVNKATYCSVGDELQITPQGRDEISLDSAKKACESEFTDVTNTEHQ 240  
 Db 181 ELGSDSEEDTVNKATYCSVGDELQITPQGRDEISLDSAKKACESEFTDVTNTEHQ 240  
 QY 241 PSNNDLNTTEKRAERHPEKYGSSVSNLHVPCGNTHASLQHNSSLTLTDRMVE 300  
 Db 241 PSNNDLNTTEKRAERHPEKYGSSVSNLHVPCGNTHASLQHNSSLTLTDRMVE 300  
 QY 301 KAEFCNKSQPGIARSQHNRMAGSKETCNDRTPTSTKRVLDLNDPLCERKEMKOKLPC 360  
 Db 301 KAEFCNKSQPGIARSQHNRMAGSKETCNDRTPTSTKRVLDLNDPLCERKEMKOKLPC 360  
 QY 361 SENPRDTEVPWITLNSIQKVENWFSNDELGSDSDSHGSESNNAKAVADYLDVLEVD 420  
 Db 361 SENPRDTEVPWITLNSIQKVENWFSNDELGSDSDSHGSESNNAKAVADYLDVLEVD 420  
 QY 421 EYSGSSEKIDLASDPHEALICKSERVHSKSVESNIEKIFGKTYRKASLPNLSHTEN 480  
 Db 421 EYSGSSEKIDLASDPHEALICKSERVHSKSVESNIEKIFGKTYRKASLPNLSHTEN 480  
 QY 481 LITGAFTVEPQIIOERPLTNKLRKRKRPSTGLHPEDFTKADLAVOKTPKEMINGNOTE 540  
 Db 481 LITGAFTVEPQIIOERPLTNKLRKRKRPSTGLHPEDFTKADLAVOKTPKEMINGNOTE 540  
 QY 541 QNGQVANTNSGHEKTKGDSIQNFKNNPPIESLEKSAFTKAPLISSISNNELELNI 600

Db 541 QNGVAMNTSGHENTKGDSONEKNPNPDESLEKESAFKRAEPISSISIMMELNI 600  
 QY HNSKAPKRNLRKRSSTRHIALELVYSRLSPNCTELOIDSCSSSEETKRRKNOMPV 660  
 Db 601 HNSKAPKRNLRKRSSTRHIALELVYSRLSPNCTELOIDSCSSSEETKRRKNOMPV 660  
 QY 661 RHRNRLOJMEGKBPAGKAKSNKPNQOTSKRHSDFPELKLTNAPGSPFKGNTSELKE 720  
 Db 661 RHRNRLOJMEGKBPAGKAKSNKPNQOTSKRHSDFPELKLTNAPGSPFKGNTSELKE 720  
 QY 721 FVNPSPREKEKLELVYKVSNNADPKDMLSGERLOTEREVSSISLVGTDTGTQ 780  
 Db 721 FVNPSPREKEKLELVYKVSNNADPKDMLSGERLOTEREVSSISLVGTDTGTQ 780  
 QY 781 ESISLLEVTIGKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTDFGFTYPLGHEVNH 840  
 Db 781 ESISLLEVTIGKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTDFGFTYPLGHEVNH 840  
 QY 841 RETSIEMEESELDQYLONTFKVSKRSQSFALFSPNGAEBECATFSAHSGSLKQSPKVT 900  
 Db 841 RETSIEMEESELDQYLONTFKVSKRSQSFALFSPNGAEBECATFSAHSGSLKQSPKVT 900  
 QY 901 FECEQKEENGKRNESNKPQYVNTAGFPVYQKDPVNMAGCSIKGSRFLCSQFRG 960  
 Db 901 FECEQKEENGKRNESNKPQYVNTAGFPVYQKDPVNMAGCSIKGSRFLCSQFRG 960  
 QY 961 NETGLTPPNKHLONPYRIPLEPIKSFYKCKCKNLLENPEEHSMSPEREMGNIP 1020  
 Db 961 NETGLTPPNKHLONPYRIPLEPIKSFYKCKCKNLLENPEEHSMSPEREMGNIP 1020  
 QY 1021 STVSTISRRNIRENVFKEAASNNINEVGSSTNEVGSSINEIGSSDENIQAELRNRPKL 1080  
 Db 1021 STVSTISRRNIRENVFKEAASNNINEVGSSTNEVGSSINEIGSSDENIQAELRNRPKL 1080  
 QY 1081 NAMLRIGVLOPEYVKOSLPSSNCKHPEIKQOEYEVYQYVNTDPSPLISDNLQOPGSS 1140  
 Db 1081 NAMLRIGVLOPEYVKOSLPSSNCKHPEIKQOEYEVYQYVNTDPSPLISDNLQOPGSS 1140  
 QY 1141 HASQVSEPPDLDLDEGETSFANDIKESSAVFSKQKELSRSPSPFTHHLAQ 1200  
 Db 1141 HASQVSEPPDLDLDEGETSFANDIKESSAVFSKQKELSRSPSPFTHHLAQ 1200  
 QY 1201 GYRRGAKKLESSEENLSEDELPFCOHLFGKVNNTPSQSTRHSTVATBCLSKNTEENL 1260  
 Db 1201 GYRRGAKKLESSEENLSEDELPFCOHLFGKVNNTPSQSTRHSTVATBCLSKNTEENL 1260  
 QY 1261 LSTKNSLNDOSNOVILAKASQEHLSERTKCSALFSSQCSLEDELTLANTNTODPFLIGS 1320  
 Db 1261 LSTKNSLNDOSNOVILAKASQEHLSERTKCSALFSSQCSLEDELTLANTNTODPFLIGS 1320  
 QY 1321 SKQMRHOSQGVGLSDKELVSDDERGTGLEENNOEEOSSMDNLSGAAAGCESETSYSE 1380  
 Db 1321 SKQMRHOSQGVGLSDKELVSDDERGTGLEENNOEEOSSMDNLSGAAAGCESETSYSE 1380  
 QY 1381 DCSGLSSQSDILTTQORDTMOHNLKLODEMALEAVLEOHSQSPNSYPSIISDSALE 1440  
 Db 1381 DCSGLSSQSDILTTQORDTMOHNLKLODEMALEAVLEOHSQSPNSYPSIISDSALE 1440  
 QY 1441 DLNPNQOSTSEKAVLTSQKSESEPIQONPGLSADFEVYADSTSKNKEPVERSSPSK 1500  
 Db 1441 DLNPNQOSTSEKAVLTSQKSESEPIQONPGLSADFEVYADSTSKNKEPVERSSPSK 1500  
 QY 1501 CPSLDRMYHSCGSLONRNYPQOEBELKVVDVECOULESGPHDUTETSYLPRDLEG 1560  
 Db 1501 CPSLDRMYHSCGSLONRNYPQOEBELKVVDVECOULESGPHDUTETSYLPRDLEG 1560  
 QY 1561 TPYLESGISLFSDDPESDPESEDRAPESARVGNIPSSSTALKVPOLKYAESAOQPAHAHT 1620  
 Db 1561 TPYLESGISLFSDDPESDPESEDRAPESARVGNIPSSSTALKVPOLKYAESAOQPAHAHT 1620  
 QY 1621 DTAGYNAEMESVREKPELASTERVNRKMSMYVSGLTPEEFMLVYKFAKHHITLTNLI 1680  
 Db 1621 DTAGYNAEMESVREKPELASTERVNRKMSMYVSGLTPEEFMLVYKFAKHHITLTNLI 1680

QY 1681 TEETHVAMKDAEFVCEFTLKTYFLGIAGKMWVSFMYWTOSIKERKMLNEHDFEVRGDV 1740  
 Db 1681 TEETHVAMKDAEFVCEFTLKTYFLGIAGKMWVSFMYWTOSIKERKMLNEHDFEVRGDV 1740  
 QY 1741 VGNRNHOGPKRARESODRKIFRGLEICCYGPFTNMPDQLEMMVYOLCGASVYELSSFTL 1800  
 Db 1741 VGNRNHOGPKRARESODRKIFRGLEICCYGPFTNMPDQLEMMVYOLCGASVYELSSFTL 1800  
 QY 1801 GTGVHPIVYVQPDWMTEDNCFHAIIGOMCEAPVYTRENVLDSVALYOCQELDTYLIPIPH 1860  
 Db 1801 GTGVHPIVYVQPDWMTEDNCFHAIIGOMCEAPVYTRENVLDSVALYOCQELDTYLIPIPH 1860  
 QY 1861 SHY 1863  
 Db 1861 SHY 1863

RESULT 7  
 US-10-022-819-2  
 ; Sequence 2, Application US/10022819  
 ; Publication No. US20030027166A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALLEN, Antoinette C. P.  
 ; LAWRENCE, Tammy  
 ; OLSEN, Sheri J.  
 ; RABIN, Mark B.  
 ; ANGELLY, Tracy S.  
 ; TITLE OF INVENTION: CODING SEQUENCE HAPLOTYPE OF THE HUMAN  
 ; BCRA1 GENE  
 ; NUMBER OF SEQUENCES: 67  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan Lewis & Bockius LLP  
 ; STREET: 1111 Pennsylvania Avenue  
 ; CITY: Washington DC  
 ; STATE: District of Columbia  
 ; COUNTRY: USA  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/022,819  
 ; FILING DATE: 22-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 09/074,452  
 ; FILING DATE: 1998-05-06  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: <Unknown>  
 ; REGISTRATION NUMBER: <Unknown>  
 ; REFERENCE/DOCKET NUMBER: 044921-5049-01-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-739-3000  
 ; TELEFAX: 202-739-3001  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1863 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: HOMO SAPIENS  
 ; STRAIN: BCRA1  
 ; HAPLOTYPE: OM14  
 ; POSITION IN GENOME:  
 ; CHROMOSOME/SEGMENT: 17

MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-022-819-2

Query Match 99.8%; Score 9627; DB 9; Length 1863;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1859; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MDLSALRVEEYQVNIAMOKIIECPICLDELKEPEVSTKCDHIFCKFCMLKLNOKKPSQ 60
DB 1 MDLSALRVEEYQVNIAMOKIIECPICLDELKEPEVSTKCDHIFCKFCMLKLNOKKPSQ 60
QY 61 CPECKNDITKRSLOESTRESQVLEELIKIICAFQDPTGLEANSYNFAKKENNSPEHLKD 120
DB 61 CPECKNDITKRSLOESTRESQVLEELIKIICAFQDPTGLEANSYNFAKKENNSPEHLKD 120
QY 121 EYVSIISMGYRRNRAKLLQSEPNPSLOETSLSVOLSNLQTVPTLTKOKIQOKTSVYI 180
DB 121 EYVSIISMGYRRNRAKLLQSEPNPSLOETSLSVOLSNLQTVPTLTKOKIQOKTSVYI 180
QY 181 ELSGDSSEDTVNKATYCSVGDDQLQITPOGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
DB 181 ELSGDSSEDTVNKATYCSVGDDQLQITPOGTREDEISLDSAKKACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTEKRAERHPEKYQSSVNLHVEPCGNTFHASLQHESSLLTKDRMNV 300
DB 241 PSNNDLNTEKRAERHPEKYQSSVNLHVEPCGNTFHASLQHESSLLTKDRMNV 300
QY 301 KAEFCCKSKOPGLARSOHNRNAGSKETCNDRTPTSTKRYDNLADPLCEKKEVKKOLPC 360
DB 301 KAEFCCKSKOPGLARSOHNRNAGSKETCNDRTPTSTKRYDNLADPLCEKKEVKKOLPC 360
QY 361 SENPRDEDPVITLNNLSIQKVNEFSSRDELLGSDSHDGESESNAKVADVLDVLEND 420
DB 361 SENPRDEDPVITLNNLSIQKVNEFSSRDELLGSDSHDGESESNAKVADVLDVLEND 420
QY 421 EYSGSSEKIDLLASDPHEALICKSEVYHSKSVESNIEDKIFGTYRKASLPMLSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSEVYHSKSVESNIEDKIFGTYRKASLPMLSHVTEN 480
QY 481 LITGAFTVEPQITQIEPRLNKLKRRRPTSGLHPEDFIKKADLAVOKTPEMINOQNOTE 540
DB 481 LITGAFTVEPQITQIEPRLNKLKRRRPTSGLHPEDFIKKADLAVOKTPEMINOQNOTE 540
QY 541 QNGQVNNITNSGHEKNTKGDSIONEKNPPIESLEKESAFKTAEPDISSISMMETELNI 600
DB 541 QNGQVNNITNSGHEKNTKGDSIONEKNPPIESLEKESAFKTAEPDISSISMMETELNI 600
QY 601 HNSKAPKRRRLRRKSTRIHALELVYSRNLSPNCTELOIDSCSSSEETIKKKKYNOMPY 660
DB 601 HNSKAPKRRRLRRKSTRIHALELVYSRNLSPNCTELOIDSCSSSEETIKKKKYNOMPY 660
QY 661 RRSRNLOLEMEKEPATGAKKSNKPNOTSKRHDSDTFPELKLTAAGSTFKCSTSELKE 720
DB 661 RRSRNLOLEMEKEPATGAKKSNKPNOTSKRHDSDTFPELKLTAAGSTFKCSTSELKE 720
QY 721 FVNPSLPREEKEDEKLETYKVSNNAEDEPKDMLSGEYVLOTERSVESSISLVPGTDYGTQ 780
DB 721 FVNPSLPREEKEDEKLETYKVSNNAEDEPKDMLSGEYVLOTERSVESSISLVPGTDYGTQ 780
QY 781 ESIISLLEVSTLTKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFRYPGLGHEVNS 840
DB 781 ESIISLLEVSTLTKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFRYPGLGHEVNS 840
QY 841 RETSIEMESELDAOYLONTFVYSKROSFALFSPNPAEBCATFSAHSLSLKKOSPKYT 900
DB 841 RETSIEMESELDAOYLONTFVYSKROSFALFSPNPAEBCATFSAHSLSLKKOSPKYT 900
QY 901 FECCEKKEENOGKNEINIKPVQTVNITAGFPVYGOAKDPYDNACSKSIKGSRCFLCSQFNG 960
DB 901 FECCEKKEENOGKNEINIKPVQTVNITAGFPVYGOAKDPYDNACSKSIKGSRCFLCSQFNG 960
QY 961 NETGLITPKKHGLLONPYRIPLPLFIKSFVTKCKKKNLLEENFEHSHMSPEREMGENIPI 1020

```

```

DB 961 NETGLITPKKHGLLONPYRIPLPLFIKSFVTKCKKKNLLEENFEHSHMSPEREMGENIPI 1020
QY 1021 STVSTISRRNIRRENVKREASSNNINEVSGSTNEVGSINIEISSDENTIOAELGRNRKPL 1080
DB 1021 STVSTISRRNIRRENVKREASSNNINEVSGSTNEVGSINIEISSDENTIOAELGRNRKPL 1080
QY 1081 NAMLRGLVLOPEYVKOSLPGSNCKHPEIKKOEVEEVQVTVNDFSPYLLISDLNEDPMKSS 1140
DB 1081 NAMLRGLVLOPEYVKOSLPGSNCKHPEIKKOEVEEVQVTVNDFSPYLLISDLNEDPMKSS 1140
QY 1141 HASQVSETPDDLLDDEIKEDTSAENDIKSSAVFSKSVOKGELSRSPPFTHTHLAQ 1200
DB 1141 HASQVSETPDDLLDDEIKEDTSAENDIKSSAVFSKSVOKGELSRSPPFTHTHLAQ 1200
QY 1201 GYRGAKKLLESSSEENLSEDEELPCFQHLFGVNNIPSOQTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLLESSSEENLSEDEELPCFQHLFGVNNIPSOQTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLINDCSNOVILAKASQEHHLSEETKCSASFSSQCSLEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLINDCSNOVILAKASQEHHLSEETKCSASFSSQCSLEDLTANTNTQDPFLIGS 1320
QY 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLGLENNOEBSMDNLSGEAAGCESETSVSE 1380
DB 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLGLENNOEBSMDNLSGEAAGCESETSVSE 1380
QY 1381 DCSGSSQSDILTTQORDMOMHNLIKLOEMLELEVLEOHOSOPSNSPSTIISSSALE 1440
DB 1381 DCSGSSQSDILTTQORDMOMHNLIKLOEMLELEVLEOHOSOPSNSPSTIISSSALE 1440
QY 1441 DLNRPQOSTSEKAVILTSOKSSEYPISONPEGLSADKFEVSAOSSTSKNKEPVERSSPSK 1500
DB 1441 DLNRPQOSTSEKAVILTSOKSSEYPISONPEGLSADKFEVSAOSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNTPSOBELIKVVDEBQLEESGPHDLTETSYLPRQDLEG 1560
DB 1501 CPSLDDRWYMHSCSGSLQNRNTPSOBELIKVVDEBQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFDSDPSPSEDPAPESARVGNIPSTSLKYPOLKVAESAGCPAAHTT 1620
DB 1561 TPYLESGISLFDSDPSPSEDPAPESARVGNIPSTSLKYPOLKVAESAGCPAAHTT 1620
QY 1621 DTAGYNAMEESYREKPELTASTERYNKRMSVVGSLPPEEFMLVYKPARKHHTLTMLI 1680
DB 1621 DTAGYNAMEESYREKPELTASTERYNKRMSVVGSLPPEEFMLVYKPARKHHTLTMLI 1680
QY 1681 TEETTHVYMKTDAEFYVCERTLKYFLGIGAGKVVSYFWVTOSIKERKMLNEHDEFVRGDV 1740
DB 1681 TEETTHVYMKTDAEFYVCERTLKYFLGIGAGKVVSYFWVTOSIKERKMLNEHDEFVRGDV 1740
QY 1741 VNGRNHQGPKRAESODRKIFRGLFICCYGPTNNPTQOLEMVOLOGASVYKELLSFTL 1800
DB 1741 VNGRNHQGPKRAESODRKIFRGLFICCYGPTNNPTQOLEMVOLOGASVYKELLSFTL 1800
QY 1801 GTGVHPDIVVQPDAMTDENGFHAIGMCEAPVYTRBMWLDVALYQCELDLYLIPQIPH 1860
DB 1801 GTGVHPDIVVQPDAMTDENGFHAIGMCEAPVYTRBMWLDVALYQCELDLYLIPQIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863

```

## RESULT 8

US-09-815-242-12713  
; Sequence 12713; Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel

```

APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12713
LENGTH: 2344
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match      3.7%; Score 357; DB 10; Length 2344;
Best Local Similarity 18.8%; Pred. No. 1,1e-10;
Matches 340; Conservative 248; Mismatches 780; Indels 444; Gaps 57;

QY 65 KNDITKRSLOE-----STRFSQLYEE-----LAKIICAFOLDTGLEVA 102
DB 745 KXEVTNMSKSDSVSTGSGTQOSVSTSKADSOASTSGSMTSTASSTKSTSVLS 804
QY 103 NSVNFAPK-----KENNS-----PEHLKDEYSIIISQMYRRRAKRLDSEENPSLOETSIS 153
DB 805 DSVASAKSLSTSESNVSSSTSTSLVNSOVSSSMGSVSKSTSLDFTSNSSSTKSES 864
QY 154 VOLSNGVRLTLTKRIQIOPKTSVYIELGSDSEPTVKNKATYCSVGDQELLQITPQGR 213
DB 865 VSTSTSDSLRT-----STLSLSVSMSTSGSLSKSQ----- 895
QY 214 DEISLDSAKKACESETDVTNTEHHOPSNNDLNTTEKRAERHPEKYOQSVSNLHVEP 273
DB 896 -SISTSTDSASTSGVSDST-----SNISTSESLSE-----SGSTSESIISIN 939
QY 274 CGTNT---HASSLOHNSSL-LITKDRMVEKAEPCKNSKQGLARSOINRMAGKETON 329
DB 940 SISNSVASTKLESGSTSLSTSDSKMSTSESLDSTSTSDSVGSLVAGSOSV-- 997
QY 330 DRRTPTEKKVDLADPLCEKREMNKQKLPCEPNPREDTVDVMTLNSISIQVKNEMFHS 389
DB 998 -----STSTSDSKMSTSESLSDMSTSGSLAASDSK-----MWSVSSM-STSGSGSTS 1044
QY 390 DELLGSDSDHDESESNKAVADVLDVLENDVEYSGS-----SEKIDLASDPHEALIC 442
DB 1045 ESLSIDSTSDSDSKS-----LSLSTSGSGSTSTSTSTSSSVRMSESGSTSGSMS 1094
QY 443 KSEPRHASKVESENIIEKIGKTYRKASLPNLSHTVENLITGAFVTEPPIQDERPLTNKL 502
DB 1095 TSGOS--DSTISTSTSTSDSKSASTASASESISQSVSTSTSGSVSTSTSLSTSN--SEKT 1151
QY 503 KRRRRTSLGHPEDFIKKADLAVOKTPEMINOCTNOTEQONGVMNTTNGSHENKTK----- 558
DB 1152 STSMGSTSLSTSESTSDST--STSDSISEAISSESTSLASNSNTSDSESSASA 1209
QY 559 --GDSIQNEKNPNPLESLEKESAFKTAPEISSISNMDEL-NINNSKAPKRLRRKS 615

```

```

DB 1210 FLSESLSESTSESLSGSTSDSTSLSDSNSESGSTSLSTNSGASISTSGSA 1269
QY 616 STRHIALELVARNLSPPCFTQLD-----SCSSSEIKKKKKYNNQMPVRRS-----RN 665
DB 1270 STSTVSES--VSTSLSTSTSLSTSLSTSLSTSGSKNSLSASKNSTSDSISTRK 1327
QY 666 LQIMEKEPATGAKKNNKPNQTSKRHRSD--TEPELKTNAPEGSFPCNSTSELAKEFVN 724
DB 1328 SESLASASTSLSGSTSESGSTSSSESKSDSTSMLSMQSTSGSVSTSESLDSTST 1387
QY 725 SLPREKEEKEKLETVKYSNNAEDPKDMLGGERVLOTERRSESSISLVPCTDYGTOESIS 784
DB 1388 SLSL-----SASNQSGVDNSASOSASTSTSTST--ESDSQSTSTSVTSGSTSGSES 1438
QY 785 LLEVSTLGAKAPPEPNKCVQCAFENPKGLIHCCSDNRNDTGFYPLGHEVNHRETS 844
DB 1439 -TSTSTSLSDSTSTSTSTSG-----STSTAS 1466
QY 845 IEMESELDAQYLONTFKVSKROSFAL-----FSPGNAEECATFSAHSGSLKKQSP 897
DB 1467 LSGSESDSQSISTSTSEKSKSTSTSLSDSTSTNSGSA--STSTLSNMSASSES-- 1522
QY 898 KYTECEQREENGKNESNIKPVQTVNITAGFPVVGQKDPVDNARCSIRKGSFCLSSQ 957
DB 1523 -----DSSSTSLSDSTSTASMQSS 1540
QY 958 FRQNETGLITPNKHGLQNPRIPLPIKSPKTKCKNMLEENFEHMSPEREMNE 1017
DB 1541 -----ESDSQSTSTSLSNS 1554
QY 1018 NIPST---VSTISRNIRENVFEKASSNINEVGSSTN-----EVGSSINEGSDEN 1067
DB 1555 QSTSTSTIRMTASTASESESTSESGSTSESTSDSTSTSLSDSQSTSRSTASGCASTS 1614
QY 1068 IQAELGRNPKPLNAMLRLGLVLOPEVYKOSLPGSNCKHPEIK--KOYELEVYQVTNTDFSP 1126
DB 1615 TSTSDRSTSTASTSTMTSTLD---SQMSLSTSTSTSVSDSTSLSDSVSDSTSDTS 1670
QY 1127 YLISDNLQPPMG-----SHASQVCSERPDLDDGELKEPT-----SFAINDIKESS 1174
DB 1671 TSTSGMSASISLSDSTSTSTSTASAEVMSASISDSQMSSEVSNSESVESNDSKMS 1730
QY 1175 AVFSKSVQKGLSRSPPTHTLQAGYRGAKKLESSEENLESDEBELPCFQHLFGKV 1234
DB 1731 GSTSVS--DSGSLSVTS-----LKSSESVESIS-----LSGQ 1763
QY 1235 NNIPQSTRTSTVATECLSKNTEENLLSKNSLINDSNQVILLAKASQEHLLSEETKCSAS 1294
DB 1764 SMDSVSTSDSSLSVSTSORSES--VSESDSLSDSKTSGSTSTSGSLSTSLSG 1822
QY 1295 LFSSQCELED---LTANTNTQDP-----FLIGSKQMRHOSQOGLSKELVSD 1344
DB 1823 ESYSESSLSLSDISMSDSTSTSDSLSGSISLSGSTSLSTSDSLSDSKSLSSQMSGS 1882
QY 1345 EERGTLLENNOEOS---MDSNLGEASGCEBETSVSEDCGSLSSQSDILTTQOQDTM 1401
DB 1883 ESTSTSVSDSQSTSTNSQFDSMSISASBDSMSTSDSSISGSNSTSTSLSTS--DSMS 1940
QY 1402 HMLIKLOEMAELEAVLEQHGOPSNSYPISLSDSALDLRNPDEQSTSEKAVLTQOKS 1461
DB 1941 GSY-----SVSTSTSLSDISGSIIVSDSSSTSTSLSDSMQOSQSTSTASAGSLSTSI 1995
QY 1462 EYPISONPEGLSADKEVANSSTST-----KNKEPVKRSFSPKCPPLDRWYM 1510
DB 1996 SLMSASACTLTNSQSTSVSTSLSTSDSISLISGQSAVESSTSDSTSTSDSESL 2055
QY 1511 HSCSGSLQNRNYPQOPELTKVVDVEEOOLESGPHDLTFETSYLPRDLEGTPLLEGISL 1570
DB 2056 -STSGSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 2109
QY 1571 FSDDPESDPEDEKAPASARY-----GNIPSTSLKVPOLKVAESAQSPAATHTTD-- 1621
DB 2110 SGSTSVSDSTSMSESNASISMSQSTSGSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 2169

```

QY 1622 -----TACY-----NAM-----EESVSRKPELLASTERY 1646  
| : : : : :  
Db 2170 STNSGSMNSIRHFTSLSTSGLSMSDSDSVISASESMASMSDSMSDST 2229  
QY 1647 NKRMNMYVSGLPFEEMLYKFKARKH-----HITLNLITFEET----- 1684  
| : : : : :  
Db 2230 SSSMSNSMSMSTSEENSI-----HPSMSMSTSGTHFTSTSTSTSESIAPNTNESOSTL 2282  
QY 1685 --THVVMKTDAE 1694  
| : : : : :  
Db 2283 SATSVSKHDAE 2294  
RESULT 9  
US-09-815-242-12996  
Sequence 12996, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes In  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12996  
LENGTH: 6281  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12996  
Query Match 3.5%; Score 341; DB 10; Length 6281;  
Best Local Similarity 16.9%; Pred. No. 3.3e-09;  
Matches 389; Conservative 347; Mismatches 808; Indels 762; Gaps 90;  
QY 7 RVEEYQVYINAMOKLLEPCLELKEPSTCDHIFCFMLKLLNKKGSGQCELCN 66  
| : : : : :  
Db 2848 QVEAAMNAYNATKALNGTQNLKAKQKHNTAID-----GLSHLTNAOK 2891  
QY 67 DITKSLQSTREFSOLVEELLKIICAFOLDTGLE-----YANSYNFAKKENNSP 115  
| : : : : :  
Db 2892 EALKQLVQGSTTVAAGQNEQ--ANNVDAMDKLRQSIADNATTKKQONTTDSQNK 2948  
QY 116 EHLKDEVSITSMGRNAKRLQSEPEPNSIQETSLVSQLSNLGTVR-TLTKQRIQO 174  
| : : : : :  
Db 2949 DAYNAAYTTAAG-----IIDQTSPTLDPVINOAGQVSTTKNALNGNELEAA 2998  
QY 175 KTSYIELGSDSSSEDTYKATYCSVGDO-----ELLQI-----TPQGT-RDEIS 217  
| : : : : :  
Db 2999 KOQASQSIG---SLDNLNNAOKQVTVDQINGAHYDEANQIKONANQNTLTMAGNLKQALTA 3055

QY 218 LQSAKKAACEFSETD-----VTNTEH----- 238  
| : : : : :  
Db 3056 DKDATKATVNFETDADQAKQAVNTAVTNAENITISKANGNATQAEVEQAIKOVNAKQAL 3115  
QY 239 -----HQP-----SNNDLNTTEKRAARHREKQOGSSVSLNHPGCGTTHASSL 283  
| : : : : :  
Db 3116 NGNANVQAHKDEATALLINSNDLNOAQKDAL-----KQOVQATTV-----AGVNVKQTA 3166  
QY 284 QHENSLLLTK-----DRM-----VEKAFCNKS-----OPGL 313  
| : : : : :  
Db 3167 QELNANMTQKQIADKEQTKADGNFVNADPRKQANAYQAAKALISATPDVYVTPSE 3226  
QY 314 ARSQHNRWAGSKETCNDRTPTSTERR-VDLNADPL-----CERKEWNRKQLPCSENP-D 366  
| : : : : :  
Db 3227 ITAALNKVTQAKNDLNGNTNLATAKQNVQHAIDQLPNILOAQORDREYSKQIQTQATLVPRVN 3286  
QY 367 TEDVPWITLNSLSQKYNENFNRSDLELSDSDSHDESE-----SNA----- 407  
| : : : : :  
Db 3287 AIQQAATVTLNDAMTQKQIANKAQIKSENYHDADTDKQAYDAVYTKAEELLKQTTNP 3346  
QY 408 -----KVADVLVLNEVDEYSGSSER-----IDLASDPHEALICKSERVH 448  
| : : : : :  
Db 3347 TMDPNTIQOALTKVNDTNQALNGNOKLADAKODAKTTTGLTDHLDNAOKQALTTQVEQAP 3406  
QY 449 SKSVESNIEDKTFGTYRKASLPLNLSHVTENLITGAFVTEPQIIQERPLT----- 499  
| : : : : :  
Db 3407 DIATVNVN-----KQNAQNLNNAAMTUL-----NNMLQDKTEFLNSINFDA 3447  
QY 500 NKIKRRKRPSTGILHEDFLTKAD-----LAVOKTPEMINQSTNTEBNGVYAN----- 547  
| : : : : :  
Db 3448 DOAKKQAYTNAVASHGILSKANGNASQTEQAMQVNEAKQALNGNDVQBAKDAK 3507  
QY 548 --ITNSGHENKTKGDSI-----ONEKNPPIE----- 572  
| : : : : :  
Db 3508 QVITNANQNLQAKQALQOVDAQTVANVTIKQIADQDLQAMTQKQIADQDKTKAN 3567  
QY 573 -----SLEKESAFK--TKAEPISSISIMEL-----ELN----- 599  
| : : : : :  
Db 3568 GNFVNADTDKQNAVNAVAHAEQIISGPRNAVDPQVQAQALQVNAQKGLNGNHNLYQ 3627  
QY 600 -----IHNKARKKKNLRKSSRTRIHAEIVVSNRSLSPNCTEL----- 639  
| : : : : :  
Db 3628 ARDNANTALDQLPNIQNPQKTAQKQVVS--HA-ELTVGNVAIKQNDALNNAMGTLKQ 3682  
QY 640 QIDSCS-----SSEPIKKKKYN-----OMPVRHSNRL 666  
| : : : : :  
Db 3683 QIQANSQVQSVDFQADQDQKQOAVYNNANAOQIANGIPTPLVTPVTQVQAVTTMQAK 3742  
QY 667 QLMGKEPATGAKKS-----NKNP-----EOTSRR----- 691  
| : : : : :  
Db 3743 DALNGDEKLAQKQKQALANLDTLRQDLPQRDARQINQAOALATVQOTKQONQNVNTA 3802  
QY 692 -----HDSQTFPELKLTNAPGSTKCSNSELKEFVNPSPREK 731  
| : : : : :  
Db 3803 MSNKLQGIANKQDVKASENYHDADQKQATVYNAVSAEGILN-----QITNPLNDEI 3857  
QY 732 EEKLETFVKSNAEDPKDMLSGERVLQTER-----SV 764  
| : : : : :  
Db 3858 TRALQVYTDKNG-----LNGEAKLTTEKQMAKDAVSGVTHLNDQKQALKQIOTSP 3910  
QY 765 ESSSISILVPGDYQOESISILEVSTLGAKT-----EPKCVSQAAPENPGLI 815  
| : : : : :  
Db 3911 ELATVNOVKQATATSLDQAMDQSLQAINDKAQTLADGNYLNDPRKQANVQAVAKAALL 3970  
QY 816 HGCSDKNRNDTEGFPYPLGHEVYNSHRET--SIEMESELDA-QYLONTFRV--SKROSF 869  
| : : : : :  
Db 3971 N--KQSGTNEVOAQVESITNEVNAKQALNDNDNLANKQDAKQOALNLTHLNDQKQSF 4028  
QY 870 -----ALFSNPGNAEECAFFSAHSGSLKQOS--PKVFECE-----QKRENO 910  
| : : : : :  
Db 4029 ESQITQAPLVYDVTITINQKQTLID-HAEELLRNSVADNQTTLASEDYHDATQARQNDYNO 4087

```

QY 911 GKNESNIKPVQTVNITAGFPVGOCKDPVNDMAKCSIKGSRFCLSSQFRGNETGLITPNK 970
D 4088 AYTAANNIINQTSPTMNPDPVANGATTQVNNKTVKALDGDENLAAAGQANN----- 4138
QY 971 HGLLONPRIPLPLPKSEVYTKCKKNLLEN-----FEHSMSPDEREMGN----- 1016
D 4139 -----RLDQDLHNNMAQKQOLOSIOSSDIAAVNGHKQTAESINTAMGNLINAID 4190
QY 1017 -ENIPSTVSTISNNINRENF-----KEASSNNINEGSSNN-----EVGSSINIEISS----- 1064
D 4191 HQAVEORGNITNDTDQATNTAVNEAAMINKQOGNANQTEVEQALTKYQVOTLOALN 4250
QY 1065 -DENIOAELGRNRPKLMALRLGLVQPEYK-----OSLPG 1100
D 4251 GDNHNLQV-AKTNTQALDILTSINDPOKTKLDQVTAATLVTAHQIEQANLNLQAMHG 4309
QY 1101 -----SNCH-----PEIKQEEVYQ-----TV 1120
D 4310 LROSIODNATKANSKYINEDPE--QONQDAVQAANNIINQOTATLNNAINQAAATV 4367
QY 1121 NTFESPYSLLSDNLEQMGSSHASOVSETP-----DILLDGE-----IKED----- 1162
D 4368 NT--TKAALGDVYKLDONDKDHAKQTVSQAHLNNAQKHMEDTLIDSETTTAKVKODLTEA 4425
QY 1163 -----TSFANDIKESSAVF-----SKSVQKGE 1185
D 4426 QALDQMLDALQOSIADKATRASSAVYVNPKNKQSYDEAVQAESTIAGLNPPTINKGN 4485
QY 1186 L-SRSPSPPTHTHHLAQYRRGAKKLESSSENNLSEDEELPCFOHLLFGKYNIPSOSTRH 1244
D 4486 VSSATQAVISSKNAIDGVERLADQKOTAGNSLNHLDTPTPAQOQALENOJINN----- 4537
QY 1245 SVATFCLSKNRENTL-----SLKNSLNCDSNOVILAKRSQGHLSSEFKCSASLESQ 1299
D 4538 ATRRGEVAQKLTBAQLNOMAEKLRNSIODQOTEGSK-----FINEDKPOKDAYQAA 4591
QY 1300 CSELEDLITANTN--TODPELLIGSSKOMRHOSESQVGLSDKELVSDDEERGTLSE- 1353
D 4592 VQNAKDLINQNTPTLDKQVEQLQAVNQAKNLHG--DQKLADKQHAVTDLNLNGL 4649
QY 1354 NNOEBSMDSNLGEASGCE-----SETSY 1378
D 4650 NNPORALSOJINNAATRGVNAOKLAEAKALDQOMALRNSIODQOTESGSKFINEKDP 4709
QY 1379 SEPC--SGLSQSDILITQOQRTMGNILKLOEMAELEAVLEQHSQP--SNRYPTISD 1435
D 4710 QKDAYQAAYONAKDLINQNTGNPTLDKSOVE-QLTQAVTTAKONLHGDKLARDQOQAVTT 4768
QY 1436 SSALEDLRNPEQSTSEKAVLTSQSSSEYPISONPEGLSADKFEVSADSSSTSKNK----- 1489
D 4769 VNALPRLNHAQOQALDINAAPTRE--VAQHVOQAT--ELDHAMELTKKRVQOVNT 4822
QY 1490 ---EPGEVSSPCKCSLSD---RWYHSCSGSIQNRNYPQSOELIKVYDVEQOLEE 1541
D 4823 DKAPQYTESATDCKRAVDAQLOAAESTIDPTNGSNANKDAVDQ---VLTKLOKEENEL 4878
QY 1542 SCPHDLTETSYLPRODEGRPYLESIGLSFSDPESDPSDRAPESARVGNIPSSSALK 1601
D 4879 NGNERVAEKQAKQOTIIDOLITLNLAD-QATATAKONIDQATKLOPIALEVDQATOLNOSMD 4937
QY 1602 VPOLKAESASQSPAHAHT--DTAGYNAMESVSRKPEL--TASTERYNKRMSAYVS-- 1655
D 4938 QLOQAANNEHANEQVTDYTOADSDKQNAKQALADAEVNLKQANACQVDAQOLNINLKK 4997
QY 1656 -GITPEEFMLVYKFAKKHITLTNLI 1680
D 4998 QALNGDERVALAKTKNGKHIDOLNAL 5023

```

RESULT 10  
US-10-171-311-4  
; Sequence 4, Application US/10171311  
; Publication No. US20030087270A1

```

; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangt
; APPLICANT: Gatti, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerst, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRN
; ORGANISM: Homo sapiens
; US-10-171-311-4

Query Match 3.5%; Score 333; DB 9; Length 3899;
Best Local Similarity 18.5%; Pred. No. 4.6e-09;
Matches 393; Conservative 337; Mismatches 775; Indels 622; Gaps 93;

QY 7 RVEVONVITAMOKILECPICLILKEPVSTKCDHIFKCKMILKLNQKGPSCPLCKN 66
D 1339 QVQELSLISSLO-----QOLKETQNYEAELHCLQKRLQAVSESTVPSLPDVS 1389
QY 67 DIRKRSLOESTRSQLVEELKTIICAFQDPTGLEVANSYFAKKENN-----SEPHL 118
D 1390 VITESDQRTMYGSCVKK-----NIDGTIEFSGEG-VKEETIVYKLEKQTOEOL 1440
QY 119 KDEVS-110SM-----GYRNRA--KRLQSEPNPSLOQTSVOLSNL 159
D 1441 EEEVAVIYMSJAPAQOTELSLISGKENTASSKQAHAVCOQEOHFNEMKLSQDQIGF 1500
QY 160 GIVRT--LRTKQRIQPKTSVYIELG-----SDSEDTVNAKATYCSVDQDELLQ- 206
D 1501 QTFETVDVKKFEKFEKPLSK-----ELGEGKEILLNSDPHDIPESKCVLTISEMFSSKD 1556
QY 207 ---ITPOGTDEL---SLDSAKKACEFSETD-----YNTTEHHQPSNNDLMTTERAA 254
D 1557 KTFIVRQSHIDELSVSSMDASRQMLNEBOLEDMRODELVQOYEHQOATELLRQAHMROM 1616
QY 255 ERHPEKYOGSSVSNLHVEPCGTN---THASSLOHENSLLITFDRMNVKAEFCNKSQOP 311
D 1617 ERQREDQE-----QLQELIKRLNRQLAQRSSIDNEN--LVSEKRVLELE-ALKOL 1666
QY 312 GLARSQHNRRAGSKETCNDRRTPTSTEKKYDLNADPLCEKREKMKOKLPCSENPRTQEDVP 371
D 1667 SL-----AGREKILCCELRNSJTOQ---NGNE--NQEVEVEQTFKEKELRKPEDVP 1713
QY 372 WITLNS--SIOGVN-----EWSRSDELVGSDSHSGSESNAKV- 409
D 1714 PELLSNERIALQANNRKLKILLEVYKTTAAVEETIGR--HVLGILD-RSSKSQSSASLI 1770
QY 410 ---ADVLDVINE-----VDEYSGS----- 425
D 1771 WRSEAEVSKSCVHEHTRTYDESIPSYSGDMPRNDINMWSKVTEGTELSQRLVRSQF 1830
QY 426 -----SEKIDLLAS-----DPH 437
D 1831 AGTEIDPENEELMLNISSRLOAAVERKLELAISSETSSQLHAKVQOTELMRESFROKQEA 1890

```





QY 67 DTRRSLOESTRPSOLVEELLKIIICAFOLDTGLEVANSYNFAKKENN-----SPEHL 118  
 Db 1390 VITESSOQRMTMYGSCCK-----NIDGTIEFSEFG-VKEETIYVLLKEQYOEO 1440  
 QY 119 KDEVS-1IOSM-----GYNRRA---KRLOSPEPNLSLOETSLSVOLSNL 159  
 Db 1441 EEEVAKYIVMSJAFAOQTELSRISGKENTASSKQAHAVCOOEOHFNEMKLSODQIGF 1500  
 QY 160 GYVRT--LRTKORLOPKTSVYIEEL-----SDSEEDVYNKATYCSVDQOELLO- 206  
 Db 1501 QTEETVDYVKEEFPKPLSK-----ELGEHKEILLNSDPPHIDPSKCVCULLTISEEMFSKO 1556  
 QY 207 ---ITPOGTREDEI---SLDSAKKACEFSETD-----VTNTEHQPSNDLVTTERKAA 254  
 Db 1557 KTFIVQSHIDEIVSVMASRQMLNEDQLEDMQOELVQYOEHQOATELLNOAHNRQM 1616  
 QY 255 ERHPEYQSGSVNLHVEPCGTN---THASSLOHESSLLITKDRANVEAEFCNKSOP 311  
 Db 1617 ERQREDOE---OLQDEIKRLNROLAORSSIDNEN---LVSEERERVILLELE---ALKOL 1666  
 QY 312 GLARSOHNRWAGSKETCNDRTPTSTKRYDLNADPLCEERKNKOKLPCSENPREDTIEDVP 371  
 Db 1667 SL-----AGBEKLDCELRNNSOTO---NGNE---NOGEVEBOQTFKEKELDRKPEDVP 1713  
 QY 372 WITLNS---SIQVYN-----EMFSRDELLGSDSDSHGESESNKV- 409  
 Db 1714 PELTNSRYALOKANNRLKILLEVYKTTAAVEETIGR--HVGILD-RSSKQSSASLI 1770  
 QY 410 ---ADVUDVINE-----VDREYSGS----- 425  
 Db 1771 WRSEAEASVSCYHEETRTVYDESISYSGSDMPRNDINMMSKYTEEGTELSORLYRSGF 1830  
 QY 426 ---SEKIDLLAS----- 437  
 Db 1831 AGTEIDPENNELMINTSSRLQAAVEKLEAISETSOLEHAQVOTELMRSESRQOEAT 1890  
 QY 438 EALICKS---ERYHKS-----VESNIEDK-IFGKTYRRKASLPLNLSHV 477  
 Db 1891 ESJLKQOELERLHEERAREQOAVELSKAEGVIDGVADEKTLFEPRIQOKTOI--IDRL 1948  
 QY 478 TENLIIGA---PYTEPOIIOBRPLTNKLRKRPTSG-----LHPEDFIKKADLAV 525  
 Db 1949 EOEELTASNRLOLEAEQOOIOERELLSROKEMAMAAGPVEQOULOETEKLMKELEEV 2008  
 QY 526 OKTPREMIINGTONEQOVNITNSGHEKTKGDSIONKKNPNPIESLESKESAPFKAE 585  
 Db 2009 QCOAEKVYRDDLOK---OYKALEIDVEEYVSXFTELEOEKNTELDLRQONOLEKOLE 2063  
 QY 586 PI-----SSSISNMELANI-----HNSKAPKK--NRLRKKSS 616  
 Db 2064 KMKRFLDEQAIDREHERDVQOELQLOKLVPRPROPISEHQTEVEEQLANHLKEKTD 2123  
 QY 617 TRHIALELVVSNLSPNCTEIOIDSCSSSEIKKKKYIOMVYRHSRNIQLMEGKE--P 674  
 Db 2124 ---KCESELLSKE---OLOORDIOERNEIEKLEFR--VRELEQALVEDRKHFG 2169  
 QY 675 ATGAKKSNKNEQTSKRHSDTPEPCLKLTNAPG-----SFTK 711  
 Db 2170 AVAKAPPLSLEVLOQAEORDAIDRKEKEITINLEQOLEOFREELKNKEVVOQLHMOLEIQ 2229  
 QY 712 CSNTSELKEFVNPSPREKEEKELE--TVKVSNAEDPKDMLTSGE--RVLOTERSYESS 768  
 Db 2230 KESTTRLOLEQENKLFKDKMEKLGIAIKESDAMSTODQVHLGKFAQIIQ--EKVEIIDQ 2288  
 QY 769 ISLVPGIDYGTOSISILE---VSTLGAKAKTEPNKCVS-----QCAEFENKGLIHGC 818  
 Db 2289 LN-----BOYVKLOOOLKITTDNKNVIEEKNELIRDLQOIECLMSDOE-----C 2332  
 QY 819 SKNRRND-----TEGFYUPLGHEVNHRSRETSIEMEESLEDAOYLONTFKSKROSPA 870  
 Db 2333 VKRRREIEIOLNEVIEKLOOEL---ANIGOKTSMANHSLSSEEDSISKHOLODVVIAEKLA 2389  
 QY 871 LFSNPGNAEBECATFSAHSGSLRKQSPKVTPECEO-----KEENOGKNESNIRPVOT 922

Db 2390 LEOOVETANE--WTF-----MKNVLKETNFKNMQLQOELFSLKRESEVVEKIQSIPENS 2442  
 QY 923 VNTIAGFPVYQOKDP-----VDNAKCSIKGSRFCLSSQFRQNETGLITPNKHGLON 976  
 Db 2443 VNVAIDLH---SKDKPELEVLVEDALKSLLENQYF--KSFENGSGSIINETRLLO- 2495  
 QY 977 PYRIPPLPIKSFVKTCKKNLE-----ENFEHSMSPREKGNENIPSTVSTIS 1027  
 Db 2496 ---LESTVSAKDELEIIOCYQOIKDMQOQOETEMLOKTIYVLOKIYEEKYAAALYSQIQ 2552  
 QY 1028 RNNIRE-----NYVKEASSNI-----NEVGSSTNEVGSSINEIGSSD----- 1065  
 Db 2553 LEAVQEVAKFCQDMQITISPERTINQNLQOLREDEIGSDISALTLRISLESQVEMHT 2612  
 QY 1066 ---ENIQAEIG--RNRCPKLAMRLGVLOPEYVYKQSLPSNCSNCHPEIKQOIEYEVQIV 1120  
 Db 2613 SLILEKQOEVIAEKVYLEKKEKLEL-----QKLEENKREKREKRRSPQOEVYL 2664  
 QY 1121 NTFPSYPLISDNLE---QPMGSHASQVSETPDDLLDDGETIKEDTSPFANDIKESSAV 1176  
 Db 2665 KT--TTLEFHSNBSGFFNLEBALRASVATKA-----ELASTYKKEKLOEEL 2711  
 QY 1177 FSKSVQKGLSRSPSPFTHTHLAGYRRGAKKLESSEENLSDEDELPFOHLFGKVN 1236  
 Db 2712 LVKETNMTSLQKDLQO--VRDHLAFA--KEKLSILEKDEDETVQSKKACMEPPIKLSKS 2769  
 QY 1237 IPSOSTRHSTVATECLSKNTENULSLKNS-----LNDGSONYIILAKASO----- 1281  
 Db 2770 IASO---TDGTLKLISSNQOTPOILVKNAGIOINLOSECSSEBVEIILISFEKIEKMQ 2824  
 QY 1282 ---EHHLSE-----ETKCSASLFSQSCSELDLANTNTOOPLIGSS--KQ 1223  
 Db 2825 ELHAELTDMESHRTISTETPLKREHVAVOLKREBCTTAAVLOCLRSKE---GSSITPE 2880  
 QY 1324 MRHOSQOVLGSLDKELVSDDERGTGLENNQEOGSDMSNLGEAASGCESET---SVSE 1380  
 Db 2881 LAHSDAQO---TREICSSDS--GSDMGOGIYLTSHQGFDIASEGGESESESATISFPK 2933  
 QY 1381 DCSGLSSQSDILITTOORTYQOHNLIKLOQEMAELEAVLEQHSOP--SNSIPSIISDSSAL 1439  
 Db 2934 KINGL-----LRAVHNEGMO--VLSTIESPYSDGEHSDIOVSEPWLEERKAYINTISSL 2986  
 QY 1440 EDLRNPEOSTSEKAVILTSOKSE-----YPISONPEGLSADKFEVSADSS 1484  
 Db 2987 KDLITKMOLOREAEVYDSSSHESFSDWRGELLALLOQVLEBSVLLAFRTIELTALGT 3046  
 QY 1485 T-----SKNKEPVERSSPSKCPSLDDRMYHSCSGSLONRYPSOEELI----- 1529  
 Db 3047 TDAVGLLNCLEQRIQOEGVYEQAMETLOKADR-----RSLSEIOLAHQOM 3094  
 QY 1530 -KVYVDEQOLESGRPHDLTETSYLPRQDLEGPYLESGLSIFSDPESPESDRAPESA 1588  
 Db 3095 GRKITLKREDESEKPSOELLEYNIOQO---SOMLEMOVELSS---MKDRATELO 3143  
 QY 1589 RVGNIPSPSTALVPOLKVAESAOSPAAHTTDPAGYNAMEE----- 1630  
 Db 3144 E---OLSSKMYVAELK--SELAOTKLELETTILKAQKHILKELEAFRLFKXKTDDEVHLL 3198  
 QY 1631 ---SVSREKPELTASTERYVKNKRMVSVSGLTPREFMLVYFARKHNI 1674  
 Db 3199 NDTLASQOKSRLEQNALMEKAKLGRSEERDEEL-----BDLKSLSLSQOQRNL 3249  
 QY 1675 TLNLTTEETHYVMTDAEFVCEBRL 1701  
 Db 3250 QI--NLLEEQOKOLNSESQKTESQRM 3275

RESULT 12  
 US-10-171-311-2  
 ; Sequence 2, Application US/10171311  
 ; Publication No. US20030087270A1  
 ; GENERAL INFORMATION:

```

: APPLICANT: Schlegel, Robert
: APPLICANT: Chen, Yan
: APPLICANT: Zhao, Xumei
: APPLICANT: Monahan, John
: APPLICANT: Kamathar, Shubhangi
: APPLICANT: Gatt, Karen
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Hoersth, Sebastian
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
: TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
: TITLE OF INVENTION: OF CERVICAL CANCER
: FILE REFERENCE: MRI-035
: CURRENT APPLICATION NUMBER: US/10/171,311
: CURRENT FILING DATE: 2002-06-12
: PRIOR APPLICATION NUMBER: US 60/298,159
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/298,155
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/335,936
: NUMBER OF SEQ ID NOS: 238
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 2
: LENGTH: 3907
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-171-311-2

```

Query Match 3.4% Score 329; DB 9; Length 3907;

Best Local Similarity 18.5% Pred. No. 7.6e-09; Matches 367; Conservative 313; Mismatches 726; Indels 576; Gaps 89;

```

QY 7 RVEYQVNVINAKILCEPICLELKEPVSTKCDHIFCKFCMLKLNOKKGPSCPLCKN 66
DB 1339 QVQESLSSIQ-----QOLKETQNEAEHICLQKRAQAVSESTVPSPSLVDVY 1389
QY 67 DITKSLSDESTREFSOLVELLITICAFODTGLEVANSINFAKKN-----SPEHL 118
DB 1390 VTTEDAQOTMYPGSCVKK-----NIDGTIEFGEGF-VKEETNIVKLEKQYOEOI 1440
QY 119 KDEVS-IIOSM-----GYRNRA-----KRLQSPENPSLOETSLSYVLSLN 159
DB 1441 EEEVAKIVYSMAIQAQOTELSRISGKENTASSKQAAVCOEOHYFNMKILSQOIGF 1500
QY 160 GTVKT-LRTKQRIOPKTSVYIELG-----SDSEEDTVNKATYCSVQDOELQ 206
DB 1501 QFFETVDVYKFEFPLSK-----ELGEHKEILLNSDPHDIPESDCVLTISEEMFSD 1556
QY 207 ---TIPQGRDEI---SLDSAKKACEFSETD-----VTNTEHHOPSNNDLTTERAA 254
DB 1557 KTFIVRQSIHDEISVSSMDASQMLNEBQLEDMROELVROYOEHQOATELLRQAHMRQ 1616
QY 255 EHHPEKQSSVSNLHVEPCGTN---THASSLOHENSLLTPKDRMNVKAEFCNKSOP 311
DB 1617 ERQREDOE-----QLOEITKRLNQLAQSSTIDEN---LYSEFERVLELE---ALKOI 1666
QY 312 GLARSQHNWAGSKETCNDRPTSTEEKYVDLADPLCEKRNKKOLPCSENPRTDYEY 371
DB 1667 SL-----AGREKLCLELRNNSQTQ---NGNE---NOGEVEQYFFKEKLEDRKEDYV 1713
QY 372 WITLNS---STOKVN-----EMFSRDELLGSDSDHGESENAKV- 409
DB 1714 PEILSNERYALQKANRLKILLEVYKTTAAVEETIGR---HYLGILD-RSSKSSQSSASLI 1770
QY 410 ---ADVLVINE-----VDEYSGSS---EKIDLLASDPHEALICKSERVHS-- 449
DB 1771 WRSEADASYKSCVHEHRTVDESTIPSYSGDMPRDINMSKYTEEGTGLSORLYRSGF 1830
QY 450 ---KSYESNIEDKIFGKYRKAKSLPNLSHV---TENLILGAVTEPOII-----QRRPLT 499
DB 1831 AGTEIDPEHEELMLNISSRLQAAYEKLELAISSETSSQLHAKVYOTELMRRESPROQANT 1890
QY 500 NKLKKRRPTSGLHPD-----FIKAD-----LAVOKT---PEMNOGT---NOTEO 541

```

```

DB 1891 ESLKQEEELRERLHESRAREQLAVELSKAEVIDGYADEKTLFEROIQEITDILDRLEQ 1950
QY 542 NGQVNNITNSGHNKTKGDSIONENKPNPIESLEKESAFKTAEP-----SSISNM 594
DB 1951 ---ELCASNLOLELEDOOIOEERE---LISROKE-AMKAGVYEQOOLQOETEKLMKE 2004
QY 595 EELNINHSKAPKNNLRKRSSTRHIALELVVSRNLSPPNCTELOIDSCSSSEIKKK 654
DB 2005 KLEVOQAKV-----RDDIQKYKALEIDVEQVS----- 2035
QY 655 YNOMPVRHSRNLOLEMGKEPATGAKSNK-----NEOT-SKRHSDTFP-----ELKL 702
DB 2036 -REFLEOEKNTLMDLRQONALEQOLEMKRFLDEQALIREHRDVFQOEOIKLEOOL 2094
QY 703 TNAPSEFTCS---NISELKEFVNPSLPREKKEKLETKVSNANADPDLMLSGRYLOT 760
DB 2095 KVPV-RFQPISEHOREVQLAN---HLKEKTDKSEILLKE-OLORDIOERMEIEKL 2149
QY 761 ERSVESSISLVGDTGTOESISLLEVSTLGAKTEPNKCVSQCAFENPKGLHGCCK 820
DB 2150 EFRVRELEQALLVSAD--TFQKVE--DRKHFGAVAKPE--LSLEVQARERDAIDRKEK 2203
QY 821 DNRNDTEGFKYPLGHEVNHRSRETSTIMESESLDAQYLONTKRVSKOSFALPSNGAAE 880
DB 2204 EITN-----LEOLOEOPRE---ELENNKEEVOQLHMOLEIOKKESTTRLO---ELEQ 2249
QY 881 ECATES-----AHSGSLKOSPKVTFE-----GEOKE-ENOGNESINIPVOTVN 924
DB 2250 ENKLFQDDEKLGALAKESDASTODQHYLFKKFQITIOEKREVELDQI-NEOYTKLOOOL 2309
QY 925 ITAGFPVQOKDKPVDNACSIKGSRFCLSSQFNGNETGLTPPKHGLONPYRIPLF 984
DB 2310 ITDNKVIEBKELLIDLETOIE---CLMSD----- 2337
QY 985 PIKSPYKTKCKNLLENFEE-----HSMSPER----- 1012
DB 2338 ---QECVKNRREELIEQLNEVIEKLOELANIGOKTSMANHSLEEDSLKHOLDV 2390
QY 1013 ---EMGENIPSTVSTISRNINENFK-----EASSNINEGVS 1049
DB 2391 VIAEKALQOYETANEEM---TFKNVILKETFMKNOLTOELFSLKRRESEVKIOS 2445
QY 1050 -STNEYGSSINIGSSDENIOAELGRN-----RGKLANMLRLVLO 1090
DB 2446 IPENSVNVAIDHLSKDPLEVLVETEDALKSLQENOTYKFSPEENGKGSITINLFRLLQLE 2505
QY 1091 PEYKOSLPFGSNCKHPEIKQYEEVQVNTDFSPYLSIDNLEQPMGSSHASQVCESTP 1150
DB 2506 STVSAKDELDTQC-YKQIKDMQEOGFETEMLOKKIVMLQKIVEKVAALVSOIOLEA- 2563
QY 1151 DDLIDGFEIKEDTSPAENDIKESSAVFSKSVQ-----KGLSRSPPFT----- 1194
DB 2564 ---VOETAKFCQDNQOTISSPERTNIONLNQLEDELGSLISALTYRISELSQV 2615
QY 1195 ---HTHLAAGYRRGAKLSESEENLSEDELPQFQHLIFGKVVNPIPSQSTRHSTVATEC 1251
DB 2616 VEMHTSLI---LEKEQVEIAEKVLEKEKKLLELOKLEBNEKKQOREKRRSPOVEY 2671
QY 1252 LSKNTEENLISLKNLSNCSNOVILAKASQEHILHSEFKCASLFPSSQCSLEDTANTN 1311
DB 2672 LKTTTE---LFHNSNEESGFNLEALRAE---SVATKABLASKEAEXLOEBELLVKE 2723
QY 1312 TQDPELIGSKQMR-HQSES-QGVALSKEVLSDDEER----- 1347
DB 2724 TMTSLQKDLQSVRHLAEAKKSLILEKEDETVEQESKAKCMPEPLPIKLSKIASOTD 2783
QY 1348 GT-GLEENNOEQSDSNLG---EASQCESETSVSDCSGLSSQ-----SDILTT 1394
DB 2784 GTLKISSNQTPQILVKNAGIQINQSCSSE-EVTEITISQFTKIEKMQELHAAELDM 2842
QY 1395 QORDTMQHLIK-----LOQEMALBAVLQHGQSPNSNPSII-SUSSALDELRLN 1444

```

```

Db      2843 ESRHISTETFLKRBHYAVOLLEECGTAKAVIQCLSKRKGSSIPETLAHSDAYQTRICIS 2902
Qy      1445 PEOSTSEKAVLTQSKSESEYIPSONPEGLSADKEFVSADSTSKNKEPVERSSPKCPSL 1504
Db      2903 SDGSGMDGOGIYLTHSGFPIASEGRG---ESESATDSFPKKIK---GLLRA----- 2949
Qy      1505 DDRWYHSGSGSLQNNRNPSPQOEELIKVYDEEQOLESFGPHDLTETSYLPRLDLEGTPLY 1564
Db      2950 ----YHN-----EGMOVLSTLSPSPSGEDHSIQOVS-----EPWL 2981
Qy      1565 ESGISLSEDDPESDPSDRAPESARVGNIPSSSALKVPOL-----KYAESASQSP----- 1614
Db      2982 EE-----RKAYINTLSTLKDLTTRKQLOREAEVYDSSQSHESFSDW 3022
Qy      1615 -----AAHNTDTAG-YNAEESYSREKPELTASTERY 1646
Db      3023 RGEILLALQOVFLERSVLLAARTELTALGTTDANGLLNCLQRIOEGOVEYQAAAMECL 3082
Qy      1647 NK 1648
Db      3083 OK 3084

RESULT 13
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Ghatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6

Query Match      3.4%; Score 329; DB 9; Length 3925;
Best Local Similarity 18.5%; Pred. No. 7.7e-09;
Matches 367; Conservative 313; Mismatches 726; Indels 576; Gaps 89;

```

```

Db      1501 QTEETVADVKKEFEKPKLSK-----ELGHEGKEILLNSNDPDPIDESKCVLTISEMFSKD 1556
Qy      207 ---ITPGTGDELT---SLDSAKKACFESETD-----YNTTEHOPSNNDLNTTERAA 254
Db      1557 KTFIVROSHIDELSVSSMDASROMLNEOEJDMROELVYOYOHOOATELROAHNRQM 1616
Qy      255 ERHPEKYQGSVSNLHVEPCGTN---THASSLOHENSLLTLTKDRMNAVEAFPCNSKOP 311
Db      1617 ERQREODE-----QLOEIKRLNROLAQRSSIDEN---LVSEBRVLELELE---ALKOL 1666
Qy      312 GLARSOHNRWAGSKETPCNDRRTPSTEXKVDLMDPLCERKEWVKOKLPCSEMPREDYDV 371
Db      1667 SL-----AGREKLCCELRNLSSTQTO---NGNE---NOGEVEEQTEKEKELDRKPDYDV 1713
Qy      372 WITLNS---SIOKVN-----EMFSRDELLGSDSDHGESFNKAV- 409
Db      1714 PETLSNERRVALQKANNRLKILLEVYKTTAAVEETIGR---HVIGILD-RSSKQSSASLI 1770
Qy      410 -----ADVLVDVINE-----VDEYSGSS---EKIDLIASDPHALICKSRVHS-- 449
Db      1771 WRSEAEASVSKCYHEETRTVDESIPSYSGSDMPRNDINMWSKYTEGETLSQRLVRSGF 1830
Qy      450 --KSVESNIEDKJFGKTYRRKASLPNLSHV---TENLIGAFYTEPOIT-----QERPLT 499
Db      1831 AGTEIDPENELMLNISRLOAAVEKLLAISFETSQLHAQYQTELMRESFROKEAT 1890
Qy      500 NKLKRRRRTPSGLHPED-----FIKAD-----LAVKT---PEININGT---NOTEQ 541
Db      1891 ESLKCOEELRERLHEESRAREQALAVELSKAGVIGYADERTLEFEROIOERTIIDLLEQ 1950
Qy      542 NGOVNNTTNGHNNKTGDSIONEKNPNPIESLEKESAFPTKAEPI-----SSSISM 594
Db      1951 --ELUCASNRLOELEAQQOIOERE---LLSQKE-AMAAEGVYQOOLLQETEKLMKE 2004
Qy      595 ELELNTHNSKAPKKNRLRRKSTRHIALELVYSRNLSPNCTELOIDSCSSSEETKRRK 654
Db      2005 KLEVOQCAEKV-----RDLQKOVKALEIDVEQVS----- 2035
Qy      655 YNMPVPHSNLQIMEKEPATGAKSKNP-----NQGT-SKRHDSTFP-----ELKL 702
Db      2036 -RTELEQEKNTLMDLRQONQALEKQLEKRRFLDQALDREHERVFOEIOKLEQOL 2094
Qy      703 TNAPGSFTKCS--NTSELKEFVNPSPREKEKELETVKYSNNAEDPKDMLSEVLOT 760
Db      2095 KVPV-RFQPISEHOTREVQOLAN---HLKERTKCSLLSKF-QLOORDIOERNEEIEKL 2149
Qy      761 ERSVSSSISLVPQDTYGTQESISLLEVSTLGRAKTEPNKCVSOCAFAFENPKGLIHGCSK 820
Db      2150 EFRVRELEQALLVASD--TFQKVE--DRKHFGAVEAKPE--LSLEVQLOABROAIDRKEK 2203
Qy      821 DNRNDTEGEKPYPLGHEVYNSRRETSIEMESELDQOYIONTFKYKSKROSFALFSPGNAEE 880
Db      2204 EITN-----LEEOLEDFRE---ELENNKEEVOQLHMOLEIOKKESTTRLO---ELEQ 2249
Qy      881 ECATFS-----AHSGLSKKOSPXYTPE-----CEQKE-ENGKNESNLIKPVOTVN 924
Db      2250 ENKLFKDDMKLGLAIKESDAMSTQDOHVLFQFPAQIOKEVEYIIDLQVYTKLQOOLK 2309
Qy      925 ITAGFPVVGQDKPVDNAKCSIKGSRFCLSSOFNGNETGLIPNKHGLQNPRIPLRF 984
Db      2310 ITTDMKYIEKKNELIRLLEQIOIE-----CLMSD----- 2337
Qy      985 PIKSPYTKCKKULLENFEE-----HSMSPR----- 1012
Db      2338 -----QECVKNRREEEIOLNEVIEKLOELANIGOKTSMNAHSLSEFADSLKHQLDV 2390
Qy      1013 -----MGMDNIPSTVSTISRNIRRENVK-----EASSSNINEVGS 1049
Db      2391 VIAEKALAEQVTTANDEM-----TEMKNVLAETNFKMNOUJOELSLKREBSVEKIQS 2445
Qy      1050 -STNEVGSSINFIGSSDENIOAELGRN-----RQPKINAMLRGLVIO 1090

```

```

Db 2446 IPENSVNVAIDHLSKDKPELEVLTEDALKSLENQTYFKSFEENGKSIINLETRLOLE 2505
QY 1091 PEVYQMSLPGSNCKHPEIKKOEVEVVOYVNTDFSYLLSDNLQOMGSSHAOVSETP 1150
Db 2506 STVSAKADDELJTOC-YKQIKDQOOGOFETEMLOKKYVNTLOKIVEEVVAALVSOIOLEA- 2563
QY 1151 DDLDDGEIKEDTSPFAENDIKESSAVFSKVQ-----KGELSRSPPT----- 1194
Db 2564 -----VQEVAKFCQDNQOTISSEPERINIQNLNLDREELQSDISALTRLISELESQV 2615
QY 1195 ---HHHLAOGYRGAKKLESSEBENLSSEDELPFCFOHLLFGKYNNIPISOSTRSTVATEC 1251
Db 2616 VEMHTSLI-----LEKEQVEIAERKNLEKEREKLELCKLEGNKKKOREKEREKSPDVEV 2671
QY 1252 LSKNTNEENLSLKNLSNDCSNQVYLAKASQEHHLSETPCSASLFSQSCSELDTLANTN 1311
Db 2672 LKTTTE---LHNSNESGFFNEELARAE-----SVATRAELASVKEKAEKLOEELLVKE 2723
QY 1312 TQDPFLIGSSKOMR-HOSES-QGVGLSDKELVSDDEER----- 1347
Db 2724 TMTVSLQKDLQVROHLAAKEKLSLEKDETEVQESKKACMFEPPLTKLSKSIASQTD 2783
QY 1348 GT-GLEENNOEQSMDSNLG---EASGCESETSVESDCSGSSQ-----SDILTT 1394
Db 2784 GTLKISSNQOTPOILYKNAGIOINLOSCSS-EVTEIISOFTKEKEMOELHAAEILDM 2842
QY 1395 QORDMOMHLIK-----LOQEMAELEAVLEQHSQSPNSYPLII-SDSSALDLRN 1444
Db 2843 ESRHSETETTLKREHVAVOLKEECGTAKAVYQCRSKSEGSIPELANSDAQOTREIC 2902
QY 1445 PEQSTSEKAVLTSOKSSEVPISQNPGLSADKFEVSADSTSKNKEPGEVSSPSKCPSL 1504
Db 2903 SDPSGDMQGGIYLTSHQGFDIASEGRG---ESESATDSFPKKIK--GLLRA----- 2949
QY 1505 DQRWIMHSSGSLQNRNYSQOELLKVVVEEQOLEESGPHDLTETSYLPRLQLETPYL 1564
Db 2950 -----VHN-----EGMQVLSLTSPSPYSDGDHSTIOQVS-----EPWL 2981
QY 1565 ESGISLFDSPDESDPADAPESARGNIPSPSTSAKVPOL-----KVAESAQSP----- 1614
Db 2982 EE-----RKAYINTISLKLDTTKMQLQOREAEVYSSQSHESFSDM 3022
QY 1615 -----AAHTTDTAG-YNAMEESVREKPELTASTERY 1646
Db 3023 RGEILLALQVLEERSVLLAARTELTALGTDDAVGLNCLQRIQOEGVEYOAMECL 3082
QY 1647 NK 1648
Db 3083 QK 3084

```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match      3.4%; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 4e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

QY 93 FQDITGLEVANSYNFAKKENNSPEHLKQEVSIIOGMYRNRA-----KRLQSE 141
Db 390 FTINT--EIGNNGNFG--ASLKADQFKYEVTLPQGVTVYVNSLTTFPNGNEDSTVLKNM 445
QY 142 PEPNLSQETSLSVOLSNLCTVTLTKTORIQPKT---SVYIELGSDSEDTVN---KAT 195
Db 446 TVNYVDONANKVFTSGGVTTARGTHKVELEFPDKSLKLSYKAVNANIDTPKNIDFNEELT 505
QY 196 YCSVGDOELQITPOGTRDEISLSAK--KAACEFSETDVTNTEHHQPSNNDLNTTEKRAA 254
Db 506 YRRASD-----IVINNAQEVTLTADPSVAIENMKKDALQOQVNSQVONSHTYTA--IA 558
QY 255 ERHPEKYQGSVSN--LHVEPC-----GTYTHASSLOHENSILLTKRRNNEKA 302
Db 559 EYKNRLKQADNITLINEDANVETANRASQAALDGLVTKLOAALIDNOQAIAELIDAKAOEKV 618
QY 303 EPCNRSKQOGLARSOHNRWAGSKETCNDPRPSTETKVKVDLANPDLCEKRE-----MNRK 357
Db 619 TAAQGSKKV-----TQDEVALVTKINNDKNAIAEINKQTTAOGYTTKDKNGIAYLDQV 674
QY 358 LPCSENPRTQEDVPWITLNSIIOKVNEWFSRDELLGSDSHGSESNARVADVLVNL 417
Db 675 ITPYVPAQAKQDI-----IQAVT---TRQOI-----KKSNAILOEKQVYAN 713
QY 418 EYDEVSGSSEKIDLASDPHEALICKSERVHKSVSNEIDKIRGKTYRKASAPLNLSHV 477
Db 714 D-----KIGKIEYKAIKIDDAATTAQVAAIKTKRAINDIQOTAPATTAKAALAEEDDEV 767
QY 478 TENLI-----IGAFVTEPOITIOERPLNKLKRRRPTSGLHPEDFIKKADLAVQKTPKIN 533
Db 768 VOAQIDQAPLNDPTTNEEVAEAIERIKNAK-----VSGVK-----AIEATTTAOD 812
QY 534 QCTNQTEQGYVANTTNSGCHENKTGDSITONEKNPNPIESLEKSAPFTKAEPISISSISN 593
Db 813 LERVKNKEESKIENTIDS---TQTKMAYNEVK-----QAATARTQOATVYSNMTNE 861
QY 594 MELELNIHNSKAPKNNRLKRSSTRIHIALELVSR-----NLS 632
Db 862 EVAEADAVEAAQKOG-----LHDIOVYVSKQEVADTKSKVLDKINAIOQOAKV 911
QY 633 PNCQT-----ELQIDSCSSSEETKKKKYQOMVYRHR---NLQMEGKEPATGA 678
Db 912 PAADTEVENAYVTRKQEOIYNSASTTEE-KQATVTELTQKQOEARNTIDAANTSDVTTA 970
QY 679 KKS-----NKPEQTSKRHSDTPPELKLITNAPGSFTKCSNTSELKEFVNPSLPREK-- 731
Db 971 KONGIAALNQOAAATTK--SDAKAEIA-----QASERKTAIAEMNDSTTEEDQAA 1020
QY 732 EKKLETIVKYSNNAEDPKDLMLSGERYLOTEBSVSSSISLVPGDYGTQSSISLFE---- 787
Db 1021 KDKVDOAVVYATANAD--IDNAAANTDVBNAKTTNEATIAITPDPANVPTAKQAIAIDKVOA 1078

```

```

QY 788 -----VSTLGRKTEPNKVCQACAFENPKGLIHGCSKDNBNDEGFKYPLGHEVNHRET 843
D 1079 QETALDANNKATTEERKAAKQOVOTEK-----TTADTAID-----GATHTNAVEA 1123
QY 844 SIEMSESLDQYLONTFKVSKROSFALFSPNPG-----NAEECATFSAHSGSLK 893
D 1124 AKNAEIKIETAIOPATTTKDNKQALATKANERKTAIAQOTDITAEIIAANAVDNAVT 1183
QY 894 KQSKVTFEECEOEKENGCK--NESNIKPV-QTV-----NTTAFPPVQCKDKP 938
D 1184 QANNINIEANSQNDVDQAKTTGESIDQVTPYKATAVTDAKNNTTA-----ATDDNG 1238
QY 939 VDNAK-----CSIKGSR-----FCLSSOFNGNETGLTPKHCGL----- 973
D 1239 VDTAKQAGKNSIGSTQPATVAKSKAKNDVDQAVTTQQAALDNTGTATEERKNAKADLYLK 1298
QY 974 -----LONPYRIPLFPK-----SFVKTRCKKNLEENFEESHM--- 1008
D 1299 AKERAYODIILNATQTTNDVTQIKQAVADVQITADTTIKQVARDKELATKAREQKALIAQT 1358
QY 1009 -----SPERE-----MGMENTPSTVSTISRNIRENVKKEASSNINEVGSSTN- 1052
D 1359 ADATTEKEQANOQVDAELTQGNQNIEMASQIDVNTAKDNAIOA-----IDPIQASTDV 1413
QY 1053 -----EVGSSINELIGSSDENIOAELGRNRP-----KLNAMLRLG-----V 1088
D 1414 KTNARAELETFEMQNKITELINNETNEEKNDIGPRAVEGLNNINNAATTGGVTTA 1473
QY 1089 LOPEVYK-OSLPSCNCKHPEIK-----KOEYEEVQVNTDPSF 1126
D 1474 KDTAVOKVQOHLANPNPKPAGKTAALDQAAADKKTQIEQTPNASQOEINDAKQVDETLNQ 1533
QY 1127 YLISDNLEQMGSHSQVCESETPDDLLDGE-----IKEDTFAENDICESAVSKSV 1181
D 1534 --AKTNIDQ-----SSTDEVDNAVKKGGKAKINAVKFSYKKDALAKITAAVNAKY 1583
QY 1182 QKSELSPSPFTHTLAQYRGCAKKLESSENL-----SSEDEBLPCFOHL----- 1229
D 1584 TEADNSNAS--TSSEIAEKQKLAELKOTADQNVNATSKDIEVQIHNLDININDYTT 1640
QY 1230 -----LFG-----KYNNIPSOSTR--HSTVATECLSKNTEENLISLKNLND-- 1269
D 1641 PTGKESATTDLYAVADQKKNNISADTNATQDEKQOAIKQOVQTALESINNNGDND 1700
QY 1270 -----CNSOYLAKASQEHHLSEETKCSASFSSQCSLEEDL 1306
D 1701 VDDALTOGKAIDAIOVDATVPRKANQVIDAKA-----EETK-----ESIDSDQL 1746
QY 1307 TANTNTQDPFLIGSSQMRHOSQGVGLSDKELVSDDEE-RGTGLE-----ENNOQE 1358
D 1747 TAEKTEALAMI--KQITDQAKQ--GITDATTTAEVAKAKQGLEAPNDIOIDSTEKQ 1800
QY 1359 QSNDSNLGEAASCESETSVEDCSGLSSOSDILTTQQRDTMOHNT-----TKIQ 1408
D 1801 KALE-----ELETALDQIEAGVNVAD-ATTEEKAEFTNALEDILSKATEDISDQ 1849
QY 1409 QEMAELEAV-----LEOHGSPNSYSPIISDSSALDLRNPBESTEKAVLTSQKSEYP 1464
D 1850 TTYAEIATYKNSALEOLKORINP---VYKKNALAIRVNVKQJEIIRKADADASAKE 1905
QY 1465 ISONPEGLSADKEFSVADSSTSKNKE-----GVERSSPSKCPSLDDRYMMSCSG 1515
D 1906 IATDTGGRFDRADKLID-KTOTNTEVALQONTTIAIEIAYQNPONPNAND-----TNSG 1959
QY 1516 S-----LQNRNPSQOEELIKVVDV----- 1534
D 1960 SDNNDATANSNANATPEENTGCPNVTESTJSTNNONDAATGETTATANSSEA 2019
QY 1535 -----EEQOLESSEGPDLTFTSYLPRODLEGPYLSGSLFSDDESQSEDRAPESAR 1589
D 2020 TDDANDKPPQANNSSADTSTNSPTMDNDVTSKPEVESTNGTGDPADEADNATPAESAT 2079

```

```

QY 1590 VGNIPSSSTALKVPOLKVAESAQSPAAAHPTTDAGYNAMEESV-----SREKPELTAS 1642
D 2080 NNNSTTATMENAP-----TESTATAPPTTASIGASSADSNDMSAVNSKQNAEVNNS 2132
QY 1643 TE 1644
D 2133 AE 2134

RESULT 15
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US200200615699A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12389

Query Match          3.4%; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 4e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

QY 93 FOLDTGLEAVANSYNFAKKENNSPEHLKDEYSIISQMGYRRA-----KRLQSE 141
D 390 FTINT--EIGNNGNFG--ASLKADQFKYEYTLRQGYTVVNSLTTFPNGNEDSYLKNM 445
QY 142 PEPNSIQETSLVSQLSNLGIVRLTKQRIQPKT---SYTILGSDSSEDIYN---KAT 195
D 446 TVNYDONANKVPTFSQGVTTARCTHTRKEVLFPPKSLKLSYKVAVANIDIPKNDIFNEKIT 505
QY 196 YCSVGQDQLQTLRPGCTRPDISLDSAK-KACGFSFENDVYNTPEHHQPSNNDLUTTEKRA 254
D 506 YRTASD-----LYINNAQPEVTLTADPFSYAVEMNKDALQOQVNSOVNSHTTAS--IA 558
QY 255 ERHPEKYQSSVSN--LHVEPC-----GTNTHASSLQHESSSLITTKDRMVERA 302
D 559 EYVKLQQAQNDILNEDANVHETANRASQAALDGLVTKLQALIDQALAEIDAKAQEKV 618
QY 303 EFCNKSQGLASQNNRNASKETENDKRTSTETEKVDLMDPLCEKE-----WNKQ 357
D 619 TAAQOSKRV---TQDEVALVYTKINNDKNNALAEINKQTTAQGVTEKDNGIAVLADQV 674

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2003, 10:13:36 ; Search time 62 seconds  
(Without alignments)  
2888.684 Million cell updates/sec

Title: US-09-734-672-4  
Perfect score: 9649  
Sequence: 1 MDLSALRVEEVQNVINAMQR.....LYOCQELDTYLPDPHSHY 1863

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: p1r1:.\*  
2: p1r2:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9642	99.9	1863	1 A58881	breast/ovarian can
2	4882	50.6	1812	2 T49350	breast/ovarian can
3	373	3.9	2447	2 T16870	hypothetical prote
4	365.5	3.8	3329	2 T42205	breast cancer susc
5	361.5	3.7	3343	2 T42207	breast cancer susc
6	358.5	3.7	5327	2 T13564	microtubule-associ
7	357.5	3.7	3329	2 T30904	breast cancer tumo
8	355	3.7	2954	2 T14156	kinesin-related pr
9	342	3.5	2271	2 F90073	hypothetical prote
10	341.5	3.5	1837	2 T41023	probable nuclear p
11	340.5	3.5	3187	2 JC5837	364K Golgi complex
12	340	3.5	2829	2 A42771	reticulocyte-bindi
13	339.5	3.5	3488	2 T34418	hypothetical prote
14	339	3.5	1957	2 T38077	hypothetical colle
15	334.5	3.5	3328	2 T30835	breast cancer tumo
16	333	3.5	6713	2 B89921	hypothetical prote
17	332	3.4	3225	2 I52300	giantin - human
18	330.5	3.4	5170	2 T15348	hypothetical prote
19	328.5	3.4	1875	2 S38173	myosin-like protei
20	327	3.4	3418	1 G02354	breast cancer tumo
21	327	3.4	3259	1 A56539	giantin - human
22	323	3.3	1658	2 S55101	hypothetical prote
23	322	3.3	4377	2 A55575	ankyrin 3, long sp
24	321.5	3.3	1738	2 T14867	interaplin - slime
25	320	3.3	65	2 G02989	BRCA1 - Rhesus mac
26	319	3.3	1410	1 A57013	early endosome ant
27	318.5	3.3	1526	2 A45605	mature-parasite-in
28	317.5	3.3	3507	2 T34513	hypothetical prote

30	317	3.3	2346	2 T13829	Tpr homolog - fru1
31	316.5	3.3	2484	2 T26216	hypothetical prote
32	316.5	3.3	2607	2 T26215	hypothetical prote
33	316	3.3	1727	2 T50073	myosin-like coiled
34	315	3.3	2481	2 D90011	fmrb protein (limp
35	314	3.3	1871	2 D96796	probable heat choc
36	313.5	3.2	1679	2 S48385	hypothetical protei
37	312.5	3.2	1790	2 S67593	transport protein
38	312	3.2	2938	2 T30249	cell proliferation
39	312	3.2	3924	2 S37431	ankyrin 2, neurona
40	311.5	3.2	2748	2 S57976	nuclear migration
41	310.5	3.2	2845	2 T49505	adenomatous polyop
42	309.5	3.2	2253	2 T30336	nuclear/mitotic ap
43	304.5	3.2	1744	2 JH0720	tanabin - African
44	301	3.1	1269	2 F84730	probable myosin he
45	301	3.1	2464	1 QRMSP1	microtubule-associ

## ALIGNMENTS

RESULT 1  
A58881  
breast/ovarian cancer susceptibility protein BRCA1 - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Oct-1998 #sequence-revision 16-Oct-1998 #text-change 18-Jun-1999  
C:Accession: A58881; A54652; I58130  
R:SKOINICK, M.H.  
submitted to GenBank, September 1994  
A:Description: Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete  
A:Reference number: A58881  
A:Accession: A58881  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1863 <SNO>  
A:Cross-references: GB:U14680; NID:g555931; PIDN:AAA73985.1; PID:g555932  
R:MIJI, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K.; Tavtigian, J.; Hattier, T.; Phelps, R.; Haugen-Strano, A.; Katcher, H.; Yakumo, K.; Gholami, O.; Norris, F.H.; Helvering, L.; Morrison, P.; Rostock, P.; Lai, M.; Barrett, J Science 266, 66-71, 1994  
A:Authors: Lewis, C.; Neuhausen, S.; Cannon-Albright, L.; Goldgar, D.; Wiseman, R.; K A:Title: A strong candidate for the breast and ovarian cancer susceptibility gene BRC A:Reference number: A54652; MUID:95025896; PMID:7545954  
A:Accession: A54652  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-444, 'D', 446-487, 'S', 489-593, 595-600, 'W', 601-1452, 'VLQ', 1456-1471, 'X', 14 A:Cross-references: GB:U14680  
R:Hosking, L.; Trowsdale, J.; Nicolai, H.; Solomon, E.; Foulkes, W.; Stamp, G.; Signe Nature Genet. 9, 343-344, 1995  
A:Title: A somatic BRCA1 mutation in an ovarian tumour.  
A:Reference number: I58130; MUID:95315980; PMID:7795636  
A:Accession: I58130  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1812-1813, 'PGQRTMASMGRCVRHLM' <HOS>  
A:Cross-references: GB:S78558; NID:g1000568; PIDN:AAB34725.1; PID:g1000569  
A>Note: frame shift mutation from an ovarian tumor  
C:Genetics:  
A:Gene: GDB:BRCA1  
A:Cross-references: GDB:126611; OMIM:113705  
A:Map position: 17q21-17q21  
C:Superfamily: breast/ovarian cancer susceptibility protein BRCA1: RING finger homolo C:Keywords: mammary gland; nucleus; ovary; tumor suppressor; zinc finger  
F:20-70/Domain: RING finger homology <RNG>  
F:24-64/Region: zinc finger C3HC4 motif  
F:607-621/Region: nuclear location signal  
F:651-664/Region: nuclear location signal  
F:679-692/Region: nuclear location signal  
F:24,27,44,47/Binding site: zinc (Cys, His, Cys, Cys) #status predicted  
F:39,41,61,64/Binding site: zinc (Cys, His, Cys, Cys) #status predicted  
Query Match 99.9% Score 9642; DB 1; Length 1863;

	Best Local Similarity	99.9%	Pred. No. 0:	Matches 1862:	Conservative	0:	Mismatches	1:	Indels	0:	Gaps	0:
QY	1	MDLSALRVEOVNVIAMOKILECPICELIKPEVSTKCDHICFKCMKLLNOKKGP	60									
Db	1	MDLSALRVEOVNVIAMOKILECPICELIKPEVSTKCDHICFKCMKLLNOKKGP	60									
QY	61	CPCLKNDITKRSLOESTRESQVLEELLKICAFOLDTGLEANSYNFAKKENSP	120									
Db	61	CPCLKNDITKRSLOESTRESQVLEELLKICAFOLDTGLEANSYNFAKKENSP	120									
QY	121	EVSITIQSMGYRNRAKRLLOSPEPNPSLOETSLSVQJSLNGTVTLTKQRIQ	180									
Db	121	EVSITIQSMGYRNRAKRLLOSPEPNPSLOETSLSVQJSLNGTVTLTKQRIQ	180									
QY	181	ELGSDSEEDTVNKATKATCSYQDOELLQITPOGTDEISLDSAKKAACEFSE	240									
Db	181	ELGSDSEEDTVNKATKATCSYQDOELLQITPOGTDEISLDSAKKAACEFSE	240									
QY	241	PSNNDLNTTEKRAERHPEKYOGSSVSNLHVEPCGNTTHASSLOHENSLL	300									
Db	241	PSNNDLNTTEKRAERHPEKYOGSSVSNLHVEPCGNTTHASSLOHENSLL	300									
QY	301	KAEPCKNSKQPLARQHNRAAGSKETCDNRTPSTEEKVYDLNADPLCEKRE	360									
Db	301	KAEPCKNSKQPLARQHNRAAGSKETCDNRTPSTEEKVYDLNADPLCEKRE	360									
QY	361	SENPRDTEVPWITLNSIIOKYNEMFSRDELLGSDSDHGESESNAKVAD	420									
Db	361	SENPRDTEVPWITLNSIIOKYNEMFSRDELLGSDSDHGESESNAKVAD	420									
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFCKTYRKASL	480									
Db	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFCKTYRKASL	480									
QY	481	LIIGAFVPEPQIIQERPLTNKLKRRKPTSGLHPEFIKADLAOKTBEIMQ	540									
Db	481	LIIGAFVPEPQIIQERPLTNKLKRRKPTSGLHPEFIKADLAOKTBEIMQ	540									
QY	541	ONGQVANTINSGHENKTGDSIONEKNPPIESLEKSAFKTAEPDISSIS	600									
Db	541	ONGQVANTINSGHENKTGDSIONEKNPPIESLEKSAFKTAEPDISSIS	600									
QY	601	HNSKAPKKNRLRRKSTRIHIALELVVSNLSPPNCTELOIDSCSSSEET	660									
Db	601	HNSKAPKKNRLRRKSTRIHIALELVVSNLSPPNCTELOIDSCSSSEET	660									
QY	661	RHSRLQLOMEGKEPATGAKKSNKPNQOTSKRHDSDTFPELKLITNAGST	720									
Db	661	RHSRLQLOMEGKEPATGAKKSNKPNQOTSKRHDSDTFPELKLITNAGST	720									
QY	721	FVNPSLPREEKKEKLETYKVSNNNAEDPKDMLSGEVLQTERSESSISL	780									
Db	721	FVNPSLPREEKKEKLETYKVSNNNAEDPKDMLSGEVLQTERSESSISL	780									
QY	781	ESISLLEVSTIGKAKTEPNKCVSQAFAENPKGLIHGCSKDNKNDTEGE	840									
Db	781	ESISLLEVSTIGKAKTEPNKCVSQAFAENPKGLIHGCSKDNKNDTEGE	840									
QY	841	RETSTEMEESLDAQYLONTFRVSKROSALFSNQNNAEBCATFSAHSGS	900									
Db	841	RETSTEMEESLDAQYLONTFRVSKROSALFSNQNNAEBCATFSAHSGS	900									
QY	901	FECBEKRNQGNESNIRKPVQTVNTAGPVGOKDPVDNAKCSIKGSSRCL	960									
Db	901	FECBEKRNQGNESNIRKPVQTVNTAGPVGOKDPVDNAKCSIKGSSRCL	960									
QY	961	NETGLITPNKHGLLONPRIPPLPKSFVKTKCKKNLLEENFEHSHMS	1020									
Db	961	NETGLITPNKHGLLONPRIPPLPKSFVKTKCKKNLLEENFEHSHMS	1020									
QY	1021	STVSTISRNNTRENFKASSSNINVEGSSINIEGSSDENIOAELGRNRP	1080									
Db	1021	STVSTISRNNTRENFKASSSNINVEGSSINIEGSSDENIOAELGRNRP	1080									

Db	1021	STVSTISRNNTRENFKASSSNINVEGSSINIEGSSDENIOAELGRNRP	1080
QY	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKOEVEEVQTVNTDFSPYLIS	1140
Db	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKOEVEEVQTVNTDFSPYLIS	1140
QY	1141	HASQVSETPDDLLDDEIKEDTSPAENDIKESSAVFSKSVOKGELSRS	1200
Db	1141	HASQVSETPDDLLDDEIKEDTSPAENDIKESSAVFSKSVOKGELSRS	1200
QY	1201	GVRGAKKLEESFENLSEDEELPCQOHLHPKVVNNIPQSTRHSVTATE	1260
Db	1201	GVRGAKKLEESFENLSEDEELPCQOHLHPKVVNNIPQSTRHSVTATE	1260
QY	1261	LSLKNLNDCSNVIILAKASOEHHLSSEFKCASLFSQCSSELEDTANT	1320
Db	1261	LSLKNLNDCSNVIILAKASOEHHLSSEFKCASLFSQCSSELEDTANT	1320
QY	1321	SKOMRHQESQGVGLSDKELVSDDEERGTLGEENNOEQSMDSNLGEA	1380
Db	1321	SKOMRHQESQGVGLSDKELVSDDEERGTLGEENNOEQSMDSNLGEA	1380
QY	1381	DCSGLSQSDILITTOQRDTMOHNLIKLOEMALEAVLEQHSQSPSN	1440
Db	1381	DCSGLSQSDILITTOQRDTMOHNLIKLOEMALEAVLEQHSQSPSN	1440
QY	1441	DLRNPQSTSEKAVLTSQKSEYPISONPEGLSADKFEVSADOSTSK	1500
Db	1441	DLRNPQSTSEKAVLTSQKSEYPISONPEGLSADKFEVSADOSTSK	1500
QY	1501	CPSLDDRMYMHSCGSLQNRNYPQOEELIKVVDVEEQULEESGPHDL	1560
Db	1501	CPSLDDRMYMHSCGSLQNRNYPQOEELIKVVDVEEQULEESGPHDL	1560
QY	1561	TPYLESIGLSFSDPDSDEDRAPBSARVGNIPSTSLAKVQPKVASE	1620
Db	1561	TPYLESIGLSFSDPDSDEDRAPBSARVGNIPSTSLAKVQPKVASE	1620
QY	1621	DTAGYNAMEESVREKPELTASTERYNKRMSVVGGLPPEEPMUYK	1680
Db	1621	DTAGYNAMEESVREKPELTASTERYNKRMSVVGGLPPEEPMUYK	1680
QY	1681	TEETHVVKTDAEFYCERLTLYFLGIAGKVVVSYFWVTOSIKERKM	1740
Db	1681	TEETHVVKTDAEFYCERLTLYFLGIAGKVVVSYFWVTOSIKERKM	1740
QY	1741	VGNRNHQPRAARESDRKFIFGLFICCGPFTNMPDQLEWVQLCGAS	1800
Db	1741	VGNRNHQPRAARESDRKFIFGLFICCGPFTNMPDQLEWVQLCGAS	1800
QY	1801	GTCVHPITVVOPDAMTENDNGFHAIGOMCEAPVYTRREWLDSVALY	1860
Db	1801	GTCVHPITVVOPDAMTENDNGFHAIGOMCEAPVYTRREWLDSVALY	1860
QY	1861	SHY 1863	
Db	1861	SHY 1863	

RESULT 2  
 149350  
 breast/ovarian cancer susceptibility protein BRCA1 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 12-Nov-1999  
 C:Accession: I49350  
 R:Harquís, S.T.; Rajan, J.V.; Wynshaw-Boris, A.; Xu, J.; Yin, G.Y.; Abel, K.J.; Weber  
 Nature Genet. 11, 17-26, 1995  
 A:Title: The developmental pattern of Brcal expression implies a role in differential  
 A:Reference number: I49350; MUID:96021028; PMID:7550308  
 A:Accession: I49350  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1812 <RES>



A:Cross-references: EMBL:U031625; NID:g988213; PIDN:AAB17114.1; PID:g988214  
 C:Genetics:  
 A:Gene: Brcal  
 C:Superfamily: breast/ovarian cancer susceptibility protein BRCAL; RING finger homology  
 C:Keywords: zinc  
 F:20-70/Domain: RING finger homology <RING>

Query Match 50.6%; Score 4882; DB 2; Length 1812;  
 Best local Similarity 56.1%; Pred. No 2.5e-209;  
 Matches 1046; Conservative 261; Mismatches 487; Indels 70; Gaps 32;

QY 1 MDLSALREVEQVNVINAMOKILECPICLEDIKEPVSTKCDHIFCKFCMLKLNQKGPQ 60  
 Db 1 MDLSAVQIOEVQVNVHAMQKILECPICLEDIKEPVSTKCDHIFCKFCMLKLNQKGPQ 60  
 QY 61 CPICKNDITKRSIQESTRFQSLVEELIKITCAFOITGLEYANISYNAKKNNSPERHLD 120  
 Db 61 CPICKNEITKRSIQESTRFQSLAEELIRIMAEELDTQMOLITNGFSFKRRNNSCERLNE 120  
 QY 121 EVSIISMGYRNRAKRLQSEPNPSLOETSLSVOLSLGTVRLTKRORIOPOKTSYVI 180  
 Db 121 EASIIISVGRNVRRLPQVEPQNALIKD-SLGVOLSLGIVSVKKNROTQPKKSVYI 179  
 QY 181 ELGSDSEEDTVNKAIVCSVDQELLQITPOGTDEISLDSAKKACEFSETDVTNETHQ 240  
 Db 180 ELDSDESEETVTKPGDCSVRDELLQITAPQEADEGLKHSABACEFSE-GIRNIEHQ 238  
 QY 241 PSMNDITMTEKRAEHEPEKYOSSVSNLHVEPCGINTHASSIQHENSILLTKDRNVE 300  
 Db 239 CS-DLNPENHATERHPEKQCSISINVCVEPCGDAAHSLAPETSLILLIEDRNAE 297  
 QY 301 KAFECNKSQOPGLARSOHNMASSKETCNDRTPTSEKRYDLNADPLCEKKNKOKLPC 360  
 Db 298 KAFECNKSQOPGLAVQOOSWMASSKETCNDROYPTGEKRYGRADSLSDKETHHQSIC 357  
 QY 361 SENRPTEDVPWITLNSLOKVNEMFSRSDLLGSDSDHGESESNAKVADVLVLENYD 420  
 Db 358 PENSGATTPDWITLNSVQKVNEMFSRTEMLTSDASARHESNAEAVVLEVSNEYD 417  
 QY 421 EYSGSEKIDLLASDHEALICKSERVHKSVSNSIEDKTFGTYRKASLPMLSHYEN 480  
 Db 418 GGFSSSRKTDVLPVPHHTLMCKSGRDFSKPVEDNITSDKTFGKSYORKGRPRPLNHYTE- 476  
 QY 481 LITGAVTEPQIOTERPLTNKLRKRRTSGLHPEDFIKKADLA-VOKTPREMNOGNOT 539  
 Db 477 -IIGTPIITEPQIOTEOEPFTKLKRR--STSLQPEDFIKKADSAQVQRTPDNINOGLDM 533  
 QY 540 EONGQVYNTINSCHENKTKGDSIQNEKNPNPIESLEKESAFKTKABPISISSINMELELN 599  
 Db 534 EPNEQAVSTTSCQENKLAGSNLQKESAHPTESLKRREPASTAGAKSINSVSDELVELN 593  
 QY 600 IHNSKAPKRNRLRKRSSTRIHALELVNSRLSPMCTELOIDSCSSSEITKAKKKNOMP 659  
 Db 594 VHSKAPKRNRLRKRSIRICALPLE-PISRNPPPCALQIDSCSSSEITKAKKKNNOOP 652  
 QY 660 VRSSRLQIMEGKEPAGKSNKPNEQTSKRSDSTPELTKLTNAPGSTKCSNTESEK 719  
 Db 653 AGHLRPELQIEDPAPADAKK-NEPNEHTRKRASDAFPEEKLMNAGLLTSCSPRKSQ 711  
 QY 720 EYVNPSPREEKEKLETYKVSNNADPKDMLSGE-RVLQTERVSESSISLVPGTDY 778  
 Db 712 GPVNPSP-PORTGEQLETRQMSDSAKELGDRVLCGPRSGKTDTRSESTSVSLVSTDYD 770  
 QY 779 TQESITLLESTGKAKTEPNKVSQCAFEFKGLHGSCKNRNDTEGFFKPLGHEVN 838  
 Db 771 TQNSVSVLDAHVRYARTGSAQCTQFVSENKELVHG-SNNAGSGTEGLKPLPLHALN 829  
 QY 839 HSRETSIEMESLEDAQYLONTKFKVSGROFALFNSPNAEECATFSAHSGSLKQSPK 898  
 Db 830 LSGE-KYEMDESLDQYLONTFOYVKRQSFALFSKPRSQKNC---AHSVSKLESPK 884  
 QY 899 VTPECQKKEENQCKNSNITKPVQTVITAGPVVGKDKPVDNAKCSIKGSRFLCSOP 958  
 Db 899 VTPECQKKEENQCKNSNITKPVQTVITAGPVVGKDKPVDNAKCSIKGSRFLCSOP 958

Db 885 VTAKGKQK-ERQGOEEFISHVQAVATVGLPYPCQEGKLAADTMCD--RGCRICLPSHHY 941  
 QY 959 RGNMTGLTPNKNKGLQNPYRIPPLPIKSFVNTCKKNLLEENFEHMSPEREMGNEN 1018  
 Db 942 RSGENGLSATGKGLSONSHFKQSVSPTRISITDNKPLTERPFEHTSTTEAVGNEN 1001  
 QY 1019 I-PSVSTISRNNIREVFEKASSNINEVGSSSTNEVGSSINIEGSSDENIQOELGNRG 1077  
 Db 1002 ILDSTHYTSLNN-RGNACEAGS-----GSINEYCSGTGDSFPQOLGRNRG 1046  
 QY 1078 PKLNLMLRLGLVLOPEYVKOSLPGSNCKNHPKIKOEYEEVQVQVNTDFSPYLISDNEQPM 1137  
 Db 1047 PKVNTVPRDLSMPQVCOQSVPVSD-KYLEIKKQEGEAVC----ADSPRLFDHLEQSM 1101  
 QY 1138 GSSHASQVCGSETPDDLLDQGEIKEDNPSFANDIKESAVFSKSVQGLSRSPTFHTH 1197  
 Db 1102 -SGKRVQVCSQETPDDLLDQGEIENHNSFGGDDIMERSAVNGSILIRESSRSPSPVTHAS 1160  
 QY 1198 LAQYRGRKAKLSESEENLSEDEELPCFOHLLFGKVNINIPQSTRSHVATCLSKNTE 1257  
 Db 1161 KSGSLHRASKLSESESDSTEDDLPCFOHLL-SRISNTP-ELTRCSSAVTORMPKAE 1218  
 QY 1258 ENILSLKNSLINDCSNOYILAKASQEHHLSEETKCSASLFSSQCSLEDLTANTQDPL 1317  
 Db 1219 GTQAPMKGSSSDCENNEVIMTEASQEHQFSDPDCSGSMFSSQHSAAQGSTANANSODSNF 1278  
 QY 1318 IGSSKQMRHQSQGVGLSDKELVSDDEERGDTGLEENNOEQSDMSNLGA-ASGCCSET 1376  
 Db 1279 IPPSKQSHQCGDEAFLSKELISDNEEMATCLEEDNDQEE--DSTIIPSEASYESET 1336  
 QY 1377 SVESDQSGLSQSDILITTOQRTDMQNLIKLOQEMAELEAVLEBOHSGQSPNSYISIDS 1436  
 Db 1337 NLSEDC-----SQSDILITTOQRATMKYLIKLOQEMAELEAVLEBOHSGQSPNSYISIDS 1392  
 QY 1437 SALEDLRNPQSTSEKAVLTSQKSEYVPIQNPBGSLADKFEVY-ADSSSTKKNKEGVER 1495  
 Db 1393 CALEDLPDLEPNMSGAAILTSKNIENPNVQNLKACDDKFOQLQHLGEPSTGDSQGMGR 1452  
 QY 1496 SSPSKCPSLDDRYMMSGSSQLQNRNVPQSOBELIKVYDVEQDLEBGPDLTETSLPR 1555  
 Db 1453 PSPFKSPPLAGSRSAHGCSRHLOKRNPSQDELLQPAGESE---ASSEPHNSTGQSLPR 1508  
 QY 1556 QDLQETPYLESIGLISLFSQD-PPESDPSSEDRAPESARVNISSSTALKVPLQKVAESAOSP 1614  
 Db 1509 RELEGTPYLGSLISLSSRPSESEK-----EPAHIGTTPASSTALKIPQGOVAFNSAA 1564  
 QY 1615 AAATHTDTAGYNAMESVSRKPELTAFTERVYKRNMSVVSGLTPPEFMLVYKFAKKHHT 1674  
 Db 1565 AGAD-----KAVGVIVSKIKPELTSSEERADRDISMVVSGLTPPEFVTVQKFAEKYRL 1617  
 QY 1675 TLINLTETEETHVYKMTDAEFVCEERTLYKFLGIAGKVVVSYFWVQSIKERMLNEHDF 1734  
 Db 1618 TLTDATITEETHVYKMTDAEFVCEERTLYKFLGIAGKVVVSYFWVQSIKERMLNEHDF 1677  
 QY 1735 EYVRGDVNRGNHOGPKRKARQSDQRIKFLGELICCYGFTMPMDQLEMMVQLOLCASVYKE 1794  
 Db 1678 EYVGDVYVGNHOGPKRKRSSRE-KLFKGLQVYCCPEFTMPMDLEERMVQLOLCASVYKE 1736  
 QY 1795 LSSFTLGTGVHPIVVVQPDAMTEDNQGFHAIGOMCEAPVYTRREVVLDSVALYQCOELDTYL 1854  
 Db 1737 LPSLTHTDQAHVAVIIVQPSAWTEDSNCPDQGLCKARLVMMQVLDLSISYRCRGIDAYL 1796  
 QY 1855 IPOI 1858  
 Db 1797 VQNI 1800

RESULT 3  
 T16870  
 hypothetical protein t13h2.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000  
 C:Accession: T16870



```

Db      1754 VSAPFTKTAVPNVAGFHQNOQOLYSDMAQAPQSTPIRTTPQPGTSGAPQAQTQSHLAQ 1813
Qy      1590 VGNIPSTSLAKVYOLK-----VAESAQSPAAMHTTDTGAYNAMEESVSRE 1635
Db      1814 LGQVNVNANQOQAAPOQOGMTTAQLOQMAQAAVNAQTAAQAQAAVAAE-AAVQAOQVAAQARA 1872
Qy      1636 KP 1637
Db      1873 AP 1874

RESULT 4
T42205
breast cancer susceptibility protein BRCA2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42205
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal,
submitted to the EMBL Data Library, February 1997
A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A:Reference number: 222073
A:Accession: T42205
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3329 <MCA>
A:Cross-references: EMBL:089652; NID:92443438; PID:92443439; PIDN:AB71377.1
A:Experimental source: strain CD1; 129Sv; ICR Swiss
A:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2

Query Match      3.8%; Score 365.5; DB 2; Length 3329;
Best Local Similarity 19.4%; Pred. No. 5,1e-08;
Matches 402; Conservative 290; Mismatches 750; Indels 635; Gaps 93;

Qy      72 SLESTFSQLVEELKLTIC-----AFQDTGLEVANSYNFAKKENNS--PEHLKD 120
Db      466 SLEDSIGKQMVSTSOACLSPIRSIKRREPLEDTGLTVFSQDGMTSTFEEHEAS 525
Qy      121 EVSTIISGMYRNRKRLLOSEPNS---LQETSLSVQLSNGTAVTLTKQIOPOKT 176
Db      526 ACGIGITLACQSRDSCIPSSVDGSPWPTLTLDTSATVK--NAGLSTLTKNRK-- 577
Qy      177 SVYIELGSDSEPTVNRKATYCSVDQDQLQITPGQTRDEISLDSAKKAECSETDV--- 233
Db      578 -----KFIY-SVSDASLQGGKKIQTHQLELTL- SAQLEASAFEPPLT 619
Qy      234 -TNTENHQPNSNDLINTTEKRAERHPEKRYOGSSVSNLHVEPCGTNTHASSLOHENSLS- 290
Db      620 FTNNNSGIPDSSD---KKRCLPNDPE-----EPSLTNSFGTATSKETSIYIHA 663
Qy      291 LITFRNRNVKAEFCNKSQGLARSOHNHMAQSKETC-NDKRTPTP---EKYDNLAA-- 344
Db      664 LISOD-LNDKCALYIEKPPYTAAREADFLCLPERECENDQKSPKVSNGEKYLVASCL 722
Qy      345 -----DPLCRKEWNKQKL-PCSENPRTDEVPV----- 372
Db      723 PSAVQLSSISFESQENLGDHNGSTLKLPLPSKPLSKADWMSREKMKMPEKLQCESC 782
Qy      373 ----ITLNSSTQKVNEMFSRSDDELGSDSHDSESNAKV-----ADVLVDLNEVDY 422
Db      783 KVNTELSKNLVEVEILCISE-----NSKTPGLLPGENITIEVASSKRSQ 827
Qy      423 SSGSEKIDLLASDPHEALICKSER---VHKSVESENIEDIK-----PKTYRKKASLP 472
Db      828 FNRNAKI-----VIOKQDQSGPFISEVAVNMNSEEPLPDGNNFAPQVTNKNKP 877
Qy      473 NLSHVTENLLIGAFVTEPQIIOERPL--TNKLKRRKRPSTGLHPEDEIKKADLAVCKTPE 530
Db      878 DLGSSVE-----LQEDLSHTQGPRLKNSPMVADEVDDAHAAQVILITDSD 924
Qy      531 MINQGTQTEONGOVNMTNSGHEKNT-KGDSIONEKNPNPISLEKESAFKTAEPRISS 589

```

```

Db      925 SLAVVHYTERKSRNNIRHOAKGTEDKDFKSSSLNMRKSDGNSDCKSEF---LDVPLN 981
Qy      590 -----SISNMELEINHNSKAPKK--NLRKRKSSR-----HIAHLEIVSRNLSP 634
Db      982 HNFQGSFRTPASNKEIKLSEHNVAKSKMFEKIEQYPRLLACIDIVNTLPLANOKKISEP 1041
Qy      635 NCTELQIDSCSSSEIEIKKKYQNPVRRSRNLQMEGEPATGAKKSNKPNEQTSKRDS 694
Db      1042 HIFDLK-----SVTTVSQSHNQSVSH-----EDT 1067
Qy      695 DTFPELKITNAPGSGFTCSNTSELKEFYVNSLPPEKEEKLITYK-----VSNMADP 747
Db      1068 DTAQOM-LSSKODPHSNLTTSOAEITELSTLIEESGSGPEFQFRKPSHIAQNTSEVP 1126
Qy      748 KDMLGSGRVLQGT-ERSVSESSISLVGTQDVGTOGESISLIEVSTLIGAKTEPNKCVSQCA 806
Db      1127 GNQNV-----VLSTASKEKKDIDLHPVDPVSGQTDHDKQFEGSA-GYKQSPHILLEDTCN 1181
Qy      807 AFENPKGLIHGCSKDNNDTE--GFKYPLGHEVHNSRETSTIEMESELDQAYLQNTFKVS 864
Db      1182 --KNT-----SCLPLNINEMFEGFCALGTKLSVSN-----A 1213
Qy      865 KRQSFALFSPNGNAEEBCA-----TFSAHSGS-----LKQSPKVTFE-----CEQK 906
Db      1214 LRKAMKLFSDIENSEEPSAKVGPFGFSSAHHDVASVFIKKONTEKSPDEKSSKQVPT 1273
Qy      907 FENO-----CKN-ESNKKPYQ-----TVNITAGPPV---QOK 935
Db      1274 LQNNIEMTTCIFVRNEPEYKIKNTKHEDSYTSQQRNMLENSDGSMSSTSG--PVYIHKGDS 1332
Qy      936 DKPVD-NAKCSIKGSRFLCSSOPRGNETGLITPNKKGILQNPYRIPPLPFKISFVTKC 994
Db      1333 DLRPADQSKC-----PESC--TOYAREENQIKINISDILCLETIMKAPEITCMKSDOKQL 1385
Qy      995 KKNLEENFEHBSNPERENGENIPSTVSTISNNIRRENYFEKASSNINEVGSSTNEY 1054
Db      1386 PSDMEQONIKREFNIS-----FOYASGKNTR--VSKESLKNKSVINFRNTEDE- 1429
Qy      1055 GSSINELGSSDENQALGRNRPKLAMLRLGYLQEPYVYQSLPGSKCKRPELK-QQY 1113
Db      1430 ---LTVISDS-----LNSKILHGNKKDMH-----TSCHKRAISIKKVF 1465
Qy      1114 EENVQVTNTPFSP-----YLISDNLQOPMSSHAS-----QVCSETPD---DLIDQGE- 1158
Db      1466 EDHFRPITYVQLPQCHPREFIEISTKEPTLISFTAGKKYKIMQESLIDKKNLFDFTQY 1525
Qy      1159 IKEDTFAEND--IKESSAVFSKSVQKGLSRSPSPFTTHLQAGYRGAKKLESSENL 1216
Db      1526 VRKTASFSGSKPLKDSKKELTLAYEKIEVY-----ASKCEMGNFV 1567
Qy      1217 SSEDDELPCFOHLLFGKVNINIPOSTRHSYVATE-----CLSKN---TEENL 1260
Db      1568 SKETEMLPQONYIMHYRQTEMKLTNSNGTSSKVOENIENNVEKNPRITCICQSSYVPTVDSA 1627
Qy      1261 LS--LKNSLNDCSNOVLAK-----ASQEH- 1284
Db      1628 LAYTDESRKTCVARESSLGSKRWLRBOGDKLGRNRTIKIECVKEHEDFAGNASTYBSHL 1687
Qy      1285 -----LSEETKCSASLFPSSQSELDL----- 1306
Db      1688 VIIRTEIDTNHVSQNVSTLSDPNVCHSYLSQSSPFCCHDMHNDSGYFLKNKIDSDVP 1747
Qy      1307 -----TAN-----TNTQ-----DPLFGSSK-- 1322
Db      1748 DMNRAGNTTISPRYSATKERNLHPQTINCYCVQKLENTNTPRHANKDVAIDPFLDSSNCK 1807
Qy      1323 -----OMRQOSEQGVGLSDKELVSDDERGTGLEENNDEOSMDSNLCLEAASGCSSET 1376
Db      1808 VGSIVLFTIASHQETE-----RTKELVTDNCKYKIV-----EQNRQSKPRPTCQTSCHKVL 1855
Qy      1377 SVSEDCGLSSQSDILTTQORDTM-OHNLIKLOQEMAELEAVLEQHGQSPNSNTPSIISD 1435

```

Db 1856 DSKDFICPSSGDCVINCINRKSDFCPHNEQILQHNOS-----MFLKRAATPPVGLT 1908  
 Qy 1436 SSALEDLRNPEQ-----STSEKAVLTSOKSSEYPIGONPEGLSADKFEVSADS 1483  
 Db 1909 WPTSKIRREPROAHNSRTYIGFSTASGKAIQVSDASLE-KARQVSEMGDAKQULSMV 1967  
 Qy 1484 STSKNEP--GVERSSPKCPSLDDRWYMHSCSGSLQ-----NRNP----- 1523  
 Db 1968 SLEGNKPHHSYKRENS-----VHSTQGVLSLPKRLPGNVNVSVEFSGSTAGK 2017  
 Qy 1524 -----SEBELIKYVDV-----EEQOLEESG-PHDLTETSYLPRDLLECTPY-LESGIS 1569  
 Db 2018 LVTVSESAALHKYKGMDEEDDLRTETHTLQHSPIPEDSKILPQPCAIRTEPEYVNSKLQ 2077  
 Qy 1570 LPSDDPESDPEDEARAVGNIPSTSTALKVPOLKVAESAQSPAANHITDAG----- 1624  
 Db 2078 KTYNDKSSLPNNK--ESSGSGTQSTIEVSLQLSQMRMODTQLVLCTKKVSHKANLIGK 2135  
 Qy 1625 YNAMESVSAREKDEL-TASTERVNRKSMNVSGLTPEEFMLVYKFAKHHITLTNLITTE 1683  
 Db 2136 EQLPQNIKVKTDKMTFSDVPYKTVNGEYKSESENYFETEAVESAKAMEDELDLDS 2195  
 Qy 1684 THVYVAKTDAEPCECTLYKFLGIAGKVVSTFWYTQSIKERKMLNEHD--FEVRGDV 1741  
 Db 2196 QTHAKCSL--FTCPQNETLNFNSRTRKRGVTVDAVQGPRIKRSLLNEFDRIIESKGS 2252  
 Qy 1742 NGRNHGPKRARES---ODRKIF---RGLICCYGPF 1772  
 Db 2253 T-----PSKSTPDGTIVKDRSLFTMHMSLEPYTCGPF 2283

# RESULT 5

T42207 breast cancer susceptibility protein BRCA2 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-Dec-1999 #sequence, revision 03-Dec-1999 #text, change 21-Jan-2000  
 C:Accession: T42207  
 R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal, R. submitted to the EMBL Data Library, February 1997  
 A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer A:Reference number: 222073  
 A:Accession: T42207  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3343 <MCA>  
 A:Cross-references: EMBL:U89653; NID:92443440; PID:92443441; PIDN:AAB71378.1  
 A:Experimental source: strain Sprague-Dawley; testes  
 C:Genetics:  
 A:Gene: BRCA2  
 C:Superfamily: breast cancer tumor suppressor BRCA2

Query Match 3.7%; Score 361.5; DB 2; Length 3343;  
 Best local similarity 19.2%; Pred. No. 7.7e-08;  
 Matches 379; Conservative 297; Mismatches 715; Indels 579; Gaps 96;

Qy 66 NDIYKRSIOESTRETSQVLEELKIIICAFQDLDTGLEVANSYFAKKENNSEHL-KDEVST 124  
 Db 1087 HMLTPOKAETELSTILE-----ESGSGEFTQFNK--PSHIAQNNTSA 1129  
 Qy 125 IQSMGRNRAKRLQSEPPNSLOETSLVSQLSNLGTVRTLRKQROPOKTVYIELGS 184  
 Db 1130 V--LGMQMAVVRASEEMKDVLDH--LPLNPSSVQIDHNKKFECLVYQKQ-----SS 1179  
 Qy 185 DSSEDPVNRKATYCSVDQELL---QITPOGTRDEISDSAKKAAACESETDVNTE----- 237  
 Db 1180 HLEEDTCNONTSCFLPIKEMEBGPGFCSALCTKLSVSEALRKAMKLPFDLEINSEEPSTK 1239  
 Qy 238 -----HH-----QPSNNDLMTTEKRA-----AERHDEKY 261  
 Db 1240 VGRPGSSCAHHDVSAVFKIKQNTDKSPDEKSSCOVTVQNNKEMTTCILVDENPENY 1299  
 Qy 262 QGSSVSNLHVEPCGTTTHASSLQHNSSLLTTRKDRANVEKAEFCNK--SKQPLASQHN 319

Db 1300 -----VKNIKODNNYTSQNRNAYKLENSDV-----SKSGTSGTVYINKGSDLP-FAAEKGN 1350  
 Qy 320 RMAGSKETC-----NDNRTPSTEEKVDL-----NAPRLDEREMNOKRIPCSENRDRED 369  
 Db 1351 KY--PESCTQYREENAQIKESVSDLTCLVYKAEETQHMSSDKQELPSKMBQNMKE 1407  
 Qy 370 VPMITLNNSTQKQVNEFRSDELLGSDSHDGESESNNAKVAADVLDVNLNEVDESGSSEKI 429  
 Db 1408 -----FNISFQ-----TASGKNIRVSKESLKNKSVNILD--QETEDLYVTSQSL 1448  
 Qy 430 DLLASDPHEALTKSER-----VHSKVSSENIEDIKFTKY--RKKASLPMLSVHTENL 481  
 Db 1449 -----NSKILGINKMKWHISCHKKSI--NFK-KVEEHHPIGVTSOLAOQYPEYE 1498  
 Qy 482 ILGAFYTER-----QIDQ-----RPLNKLKRRRPTSLGHP-----E 515  
 Db 1499 I--ESTKEPPLLSFTASGKRYAKIMQESLDKYNLPDEQYVYRKTTNQHODESKPLKORE 1556  
 Qy 516 DF-----IKRADLAVOKTPREMINQNTQTE---QNG-----QVANNITNSGHEKTKGCD 560  
 Db 1557 DYKERLLTAETKELEVYASCKEEMQNFVSKQTEMLPQNDNHYRQTEMLTNSGSPKHYG- 1615  
 Qy 561 SION--EKNPPIESLEKESAEKTKAEPIS-----SSISNHELELINHNSKA 605  
 Db 1616 NENKTEKKNR-ICCIQSSYFVTEEDSALACYGDSRKTCYGESSLSKGMKMLRQSDKL 1674  
 Qy 606 PKNRRLRKSTRH-----IHAEIVVSRNLSPPNCTELQIDSCSSSEIKKKKYNQ 657  
 Db 1675 GTRNTIETIQCVKTHTEDPAGNALYEHSLVYIKI-----EIDTSVSEMQASTLSD 1725  
 Qy 658 MPVHRSNLDLMEGKEPATGAKKSNKPNQTSKRHSDTPPELKLTMAGSFTKCSNTSE 717  
 Db 1726 PNVCHS---YLSHSFCHDDHNDGXYFLKDKIDSDVQPMKNTGNAIPKISATKE 1781  
 Qy 718 LKEFVNSLPREKEKLETVKYSNNAEDPKDLMSGEVLOTESVSSISLVPGRDY 777  
 Db 1782 IK--LHQYVNECVQKLET---NASPYAKNNIAIDAMLDLRCKRGSPVFT--TTH 1832  
 Qy 778 GTOESISLEVTSLGKAK-----TEPKVCVSQC-AAEENPKGLIHGCSKN--RNDTE 827  
 Db 1833 -SOETVPMKRIEFTDNCISKIYEQNRESKPRTQCSKALKALDNEDFICPSSGDCVINCSPM 1891  
 Qy 828 GFRYPLGHEY-NHSRETS-----IMESELDQYLOMTFYKSKRQSFALFS--- 873  
 Db 1892 AIFYPOSEOILQHNQSVGLKKAATPPVSLFETWDTCSIRGSPQEVHNPRTYGFSTASG 1951  
 Qy 874 -----NPGNAEECAFFSAHSGSLKQSPKVFEECQKRENGKKNESN-----K 918  
 Db 1952 KAVQVSDASLEKARQVSEIDGAKOLASVLSLEGNEKSHSVKRESSVHNTGVLSLR 2011  
 Qy 919 PVQTVNIT---AGFPVVGQDKPV-DNAKCSIKGSRFCLSSQFRGNETGLITPNKHL 973  
 Db 2012 KTLRGVSSVSPVSGFTAGKLVTVSESLAHVKG-----MLEERD-----LITETHTL 2060  
 Qy 974 LQNP-----YRPLPPIKSFVYTKCKKNLLENEEHSMSPEREMENIIPSTVSTIR 1028  
 Db 2061 QHSPTPEVDYSKIPP-----QCLSESRTPXSVS-----SKLOKTYNDKSR 2100  
 Qy 1029 NNIRENVKEAASSNINEVGSSTN---EVGSINIEIGSSDNDIADLGRNRGRNLAML 1084  
 Db 2101 -----SPSNKESGSSGNTQSLVSPQLSQEMERKQET--OSVLGTQVSOQRKHTL 2148  
 Qy 1085 RLGLVLOPEVYKQSLPSSNCKHPEIKQOEYEAVQVYVNTDFSPYLLISDNLQDPMGSHASQ 1144  
 Db 2149 E-----KKQNL-P-QNRIKISNKKMETPSSDVSMKTN-----VGE 2179  
 Qy 1145 VCSPTDILLDGEIKEDTSFAEND-IKESSAVFSK-----SVQKGLSRSPPTHTHL 1198  
 Db 2180 YSKPEPENYFEFAVIAIAKAFMEDELDLDSQTHAKCSLFAQPOREALINS-----PFRK 2234  
 Qy 1199 AGQYRGAKKLSSSENLSSDEDELPCQPHLLFGKVNIN--PSQSTRHSIVATECLSKYTE 1257  
 Db 2235 RGMAGVAVGQPPIKRSLNE-----FDRIIESKGSILTTPSKSTPDGTIKDRRLF----- 2284

QY 1258 ENLLSLKNSLNCDSNOVILAKASOEHLIS-EEFKCSALFSSOCSELEDLTANTQDPF 1316  
 Db 2285 -----THHMSLEPYTCGPF-----CSSSE-----RQETOSPH 2311  
 QY 1317 LISSSKOMR---HOSSESQVGLSD-----KEIVSDEER---GTGLEENNQ 1356  
 Db 2312 VTSPAOGLQSKHRSRAVGAOKSSNPTVSALNSEFTRHSHVSDKSTKVFPPEKVRSH 2371  
 QY 1357 EEOGSMOS-NIGEAASGCESTVSED-----CSGLSSQSDIILTQOORDMOH 1402  
 Db 2372 RDEFPDKNVNLBEGKNOKSADGVSEDOGNSDFPOFNKDLMSLOMARDLDIRKKNERH 2431  
 QY 1403 NLIKLOEAMLEBAVLFEHOSOPSNSYPSITISDSALEDLRNEQO---STSEKAVLTQ 1458  
 Db 2432 HLCQPSGL-----YLTKSTLPRISLOAAVGS--VPSACSPKQLYMTGVSACISVNS 2484  
 QY 1459 KSESEY---PISON--PEGISADKFEVSADSS---TSKNKEPGEVRSRSPKC--PSLDD- 1506  
 Db 2485 KMEYFOFAIEDHFGKCALCAGKGFRLADGWLIPSDGKAGKEEFYRALCDTPGVDPKL 2544  
 QY 1507 -----RWYHSCSG-----SLONRNPISOELIKV---VDEEQOLEESGPHD 1546  
 Db 2545 ISSVWVSNHYRMLVWKLAAMEFAPPEFANRCLNPEKVLQLYRYDVEIDNSSRSALKK 2604  
 QY 1547 LTE-----TSLPRLDLEGTPLYLESGLSFSDDESDPESEDRAPEARVGNL----- 1593  
 Db 2605 ILERDPTAAKTLYLCVSDI-----ISLSTNVSTSGSKASSSEDSKKNVDITELTDGW 2655  
 QY 1594 -----PSSTSALKVP---OLKVAESAOSP 1614  
 Db 2656 YAVKAQIDPLPLALVKSGLTVGOKITTOGAEIVGSDACAPLEAPDLSLKLKISANSTRP 2715  
 QY 1615 AAMHT-----TDTAGYNAMEESYRREP---ELTSTERY--NKR 1649  
 Db 2716 ARHNSKIGFHDPRPPLPLSLFSDGNGVCDVLYQRYVPLQWVEKTVSGSYIFERNR 2775  
 QY 1650 MSWVSGLPPEEFMLVYFKARKHHITLTLITEETHVVKMTDAEFYCER 1699  
 Db 2776 -----EEBEKALRFEAQOKKLEALFTK--VHTELKEHEEDLAQR 2813

RESULT 6  
 T13564  
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
 N:Alternate names: hypothetical protein EG:49E4.1  
 C:Species: Drosophila melanogaster  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13564  
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
 submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
 A:Reference number: Z17689  
 A:Accession: T13564  
 A>Status: preliminary; translated from GB/EMBL/DDBT  
 A:Molecule type: DNA  
 A:Residues: 1-5327 <SP>  
 A:Cross-references: EMBL:AL031128; PTDN:CAZ20006.1  
 C:Genetics:  
 A:Cross-references: FlyBase:FBgn0025392  
 A:Intons: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
 A:Note: EG:49E4.1  
 C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.7%; Score 358.5; DB 2; Length 5327;  
 Best Local Similarity 17.9%; Pred. No. 2e-07;  
 Matches 362; Conservative 308; Mismatches 822; Indels 525; Gaps 75;

QY 31 IKPVPSTKCDHIFCKFQMLKLNOKR-----GPSQCLCKNDITKRLQES-TRF 79  
 Db 2355 VAEIVSSPPIEAMFESKIEVEKSSIALSLQSGSGKLOTDBSPVDVAEGDSSHAVASY 2414  
 QY 80 SOLVEELLKITICAFQDLDGTGLEVANSYNFAKKENNSPEHLKDEVSITQSMGRRNRRRLQ 139

Db 2415 STVTPTLTKPAELAQIQAATVSSPLDEALRTBPAPHI-----SRADSPAE 2461  
 QY 140 SEPPNPLOQETSLSVQSLNIGTVTLTKRIQRIQKTSYVIELGSDSSEDTVKATKATCSV 199  
 Db 2462 CASEEIASODKSPVLYLKESRRPAMVAESKDDAAQKSSVE-DLKSVASTEISRP--ASA 2518  
 QY 200 GDOELLQITPOGTRDEISLSAKKACEF-----SETDVNTEHHOP---SNN 244  
 Db 2519 GETAASSPI-EEARKDFAEFEQAEKAVPLITELKGNLPTLSSPVDVAHAS-VQPAELSKV 2576  
 QY 245 DLNTEKRAAEHRPEKYOSSVSNLHVEPCGTNTHASSLOHENSLLTLTKDRNVAEAF 304  
 Db 2577 DIEKTASSPIDEAPKSLTIGSPAEREPESPASADAEE-----SVEKSK- 2620  
 QY 305 CNKSKQGLARSOHNRNRAKSKETCNDRRTSTETKVLNADPLICEKEMKOKLPCSENP 364  
 Db 2621 -DASRPVSVEST-----KADSTKGDISPSPEVLEGPDKDVEKSESR--PPSVSA 2670  
 QY 365 RDTEDVPMITLNSIOLKVNEMFSKSDLLGSDSHODESESNKAVADVLDVNEVDEYSG 424  
 Db 2671 SITGD-----STKDVSPASVSEVKDEHDKAESRSTIAKVESYIDEAGKSDSKSS 2722  
 QY 425 SSEKIDLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTENLLIG 484  
 Db 2723 SODS-----QKDEKSTLASKASRESVESKDDA-----EKSESRP----- 2760  
 QY 485 APTPEOIIIDERPILTNLKKRRPTSGLHPEDPLKAKDLAVOKTPEMINGTNOTEONGQ 544  
 Db 2761 -----ESVIASGEVPPRESK-----SPLDKDTSRPSVSEVYAE-----DEKSEQSR 2805  
 QY 545 VMNITNGHGNKTKRGDSIONENKPNPLESLEKESAFYTKAEPISSISNMELINHSK 604  
 Db 2806 RESVAEYKADTKKDGSOASRPSSVDELKDD--DEKESRQSTIGSHKAMSTGDE 2863  
 QY 605 APRKNRLRRKSSRTHIALELVYSRNL---SPNCTELOIDSCSS--BEIKKKYNNOM 658  
 Db 2864 SPMDKADKSKESPRESVAESIKHEMTKDESPIGSR--DSVAESIKDITKGEKSP 2920  
 QY 659 PVKHSRLQIMEK---EPATGAKSKNPNQOTSKRIDSTFPELKITNAPGFTKCSNT 715  
 Db 2921 PSKEVSRPESVVSIDKEKASRESVAESVKESSKDATSAPPKHSRP----- 2971  
 QY 716 SELKEFYNPLPREKEEKELETYKVSNNADDPDLMSGRVLOTESVSSSISLVPQT 775  
 Db 2972 -----ESVLGSLKDEGDKTTRRVSVADSTIDEKSLVSOE-ASRPSEASLSLDAAPSQ 3026  
 QY 776 DYGTQESISLLEVSTLAKTEPNKCVSOCAEENPKGLIHGCSKDNDRNDEGFKYPLGH 835  
 Db 3027 ETSRPEEVT--EEVKDOKSPVASKASRPASVAEN-----AKOSADES--KEGRPE 3073  
 QY 836 EVNHSRSTSIEMESLDAO-YLQNTFKVAKROSA-----LNSPQNAEE----- 880  
 Db 3074 SLPOKAGSTIKDKSPLASKDEAKSEESRRESVAEOPLVSKESVSRPASVAESVKDEA 3133  
 QY 881 -----ECATFSAGSGSLKOSPVYTFECBEKKNOCNENINIPVOTVNTAG 928  
 Db 3134 EKSKESPLMSKASRAPASVAGSVKDEAK-----SKESRRESVAEKSPLSKEKSRP 3187  
 QY 929 FPVVGOKDKRPVDAKCSIK---GGSRPCLSSQFRGNETGLITPNKGLLONPYRIPLFP 985  
 Db 3188 ASVAESYKDEADKSKESRRESGAESPLASKESRAPAS----- 3226  
 QY 986 IKSPVTKCKKNLLE---ENFEHSMSPPEMKNENIPSTVYITSRNINIEVFKKSSS 1042  
 Db 3227 VASISIDKAEKSKESRRESVAESKSPLSKEASR--PISVA-----KSYVDEAKSKES 3279  
 QY 1043 NINEVGSSTNEVGSSTINEIGSSDENIOAELGRNRGPKLNLMLGLVLO--PEVYKO-SLP 1099  
 Db 3280 SRDSVAEKSLAKESRAPASVAESVQDEAKSK---EESKRESVAEKSPLAYKESRSP 3335  
 QY 1100 GSNCK--HPEIKKQEEVYQVNTDFSPYL-----ISDNLEQ----- 1135

```

Db 3336 ASVAESIKDEAKESKESRESRY-AEKSPPLASKASRPTSVASVYKDEAEKSESSRDS 3394
QY 1136 -----PMGSSHAOCSETPDDLDLDGELKEDTSFANIKESAVESKSVQKELSRSP 1190
Db 3395 VAESPPLASKESRSPAS-VAESYQDEAEKSESRRES-VAESPLASK-----EASRPA 3447
QY 1191 SPFTHTHLAQYRGAKKLESSEEN-----SSDEDELPC-FOHLLFGKVNINPS 1239
Db 3448 S-----VAESYKDAEK--SKEESRRESVAESKPLASKASRPAVASVYKDEAEKSK 3499
QY 1240 OSTRSTVATECLSKTEENLLSLKNSLDNCSNQVTLAKASQHHLLSEETKCSATLPSQ 1299
Db 3500 ESRRSVAEKSPPLPSKASRPTSVASVYD-----EAESKESRRESVAESKSLASK 3553
QY 1300 CSELDLANTNTODPFLIGSSKOMHOSQGVGLSDE----- 1339
Db 3554 ASRPASVASVYDEAE--KSKEESRRESVAEKSPPLASKASRPAVASVYKDEAEKSK 3610
QY 1340 -----LVSDDERGTGLENNQEQSMDSNLGEAASGCESETSVSDCSGLSS 1387
Db 3611 VSRRESVAEKSPPLPSKASRPTSVASVYKDEADKSKESRREGCAKSPPLASKASRPTS 3670
QY 1388 OSDILITQORDTMQHNLLKQEBMAELVLEQHSQPSNSTPSTISDSALEDLRNPQ 1447
Db 3671 VAESYKDETEKESKEES--RRESVTEKSPPLPSKASRPT-SVAESYKDEAEKSESRRE 3726
QY 1448 STSEKAVLTSOKS-----EYPSQNPGLSADKEFY----- 1479
Db 3727 SVAESKSPPLASKESRSPASVAESIKDEAEETKQESRRESPESSKASIKGDOSSLSKET 3786
QY 1480 -----SADSTSKNKEPVERSSPKCP-----SLDDRMYHSCSGSLQNRVPSQ 1525
Db 3787 SRPDSVSVSVKDETEKPESSAIDKQSVASRPESVAVASANDKSPPLASRPESVADKSPDS 3846
QY 1536 EELIKVYVEEQOLE--ESGPHDLETSTLPRODLEG-----TPYESGISLFSDD----- 1574
Db 3847 KEASRSLVAETASSPLEEGPSIADLS-LP-LNLGEAKGKLPSTLSPIDVAEGGFLEY 3904
QY 1575 -----PE-----SDPSDRAPESARVGNIPSSSALVPOLKVAESQSPAATTTAGY 1625
Db 3905 KAESSPRPVLSPKPAEFSQPDGHTASTYVDEASPLLEIEVYO-----HTTGCVGA 3957
QY 1626 NA-----MEESVSR 1634
Db 3958 TGATAETDLDLTETKSETVTKQSETTLPETLTSKVESKVEVLESSVKQVEEKVQTSVQ 4017
QY 1635 EKPELTASTERYNKRMAMVYSGLTPEEFL----- 1666
Db 4018 AETVTYDLSLEQLTKSS--EQLTEIKSVLDINISNVTNLFSTAVETIEKKYQDVTEKYI 4074
QY 1667 KFARKH--HITLNLITEET-----THVMKTDAEF--VCERTLKY 1703
Db 4075 EKATHVSEHVTYTTGESSSTETQEKSSLDLGTFSLEIRETHITTVGSPETVITICEDEP- 4133
QY 1704 FLGIAGKMWVSYFWYTOISIKERKMLNHDFFVRCGV 1740
Db 4134 -----VLHDIKEED--EHRFSPSPDV 4153

```

## RESULT 7

T30904 breast cancer tumor suppressor Brca2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T30904

R:Connot, F.; Smith, A.; Wooster, R.; Stratton, M.; Dixon, A.; Campbell, E.; Tait, T.M.;

Hum. Mol. Genet. 6, 291-300, 1997

A:Title: Cloning, chromosomal mapping and expression pattern of the mouse Brca2 gene.

A:Reference number: 220931; MUID:97217789; PMID:9063750

A:Accession: T30904

A&gt;Status: Preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-3329 &lt;CON&gt;

A:Cross-references: EMBL:U82270; NID:g1854950; PID:g1854951; PIDN:AA848306.1  
 C:Genetics:  
 A:Gene: Brca2  
 A:Map position: 5  
 C:Superfamily: Breast cancer tumor suppressor Brca2  
 C:Keywords: tumor suppressor

Query Match 3.7%; Score 357.5; DB 2: Length 3329;  
 Best Local Similarity 19.2%; Pred. No. 1.2e-07;  
 Matches 401; Conservative 292; Mismatches 742; Indels 649; Gaps 94;

```

QY 72 SLQESTRFSQLVEELKIC-----AFQDTGLEVANSYNFAKKENS--PEHLKD 120
Db 466 SLEDSTIAGQWVSRTSOACLSPTIRKSTFFKREPLDELGTGVFSDMSNTSTFEHBS 525
QY 121 EVSIISMGYRNRAKLLQSEPNP--LOETSLVSQSLGVRVLRKQROPOKT 176
Db 526 ACGLGILTACSOERDSCIPSSVDGSMPTLTDATYK-FNAGLITLKNKR----- 577
QY 177 SVYIELGSDSEDTYKATYCSVGDDELQTPQGTRODEISLSAKKACESETDV--- 233
Db 578 -----KFTY-SYSDASLQKKKLOTHROLETLN-SQOLENSAFEVLT 619
QY 234 -TTEHHQPSNNDLNTTEKRAERHPEKYQGSVSNLHVEPGCTHTHASSLOHENSLS- 290
Db 620 FTNVNSGIPDSSD-----KKRCLPNDE-----EPSTNSFGATNSKEISYIHA 663
QY 291 LITKDMNVEKAFCKSKQOPGLANSQHNKMGSETC-NDRRTST--EKKVDLNA-- 344
Db 664 LISQD-LNDKEAIVIEEKQPYTARADFLCLPRTGENDKSPKSVNGREKVLVSACL 722
QY 345 -----DPLCKERKMNOKL-PCSENPRTQEDVPM----- 372
Db 723 PSAVOLSSITSPSQENPPLDHNCTSLTKLTPSKPLSKADNVSNKKMKPEKLOCESC 782
QY 373 ---ITLNSIQYVNEFSRDELGSDSDHSESESNKV-----ADVLVINEVDEY 422
Db 783 KNVIELSKNILEVNEICILSE-----NSKTPGLPGENIIEVASSMSKQ 827
QY 423 SGSSSEKIDLLASPHALCKSER--VHSKSVESNIEDK-----FGRTYRKKAALP 472
Db 828 FQONAKI-----VIQDKQSPFISEVAVNMNSEELPPDSGNPFAQVTKCKMP 877
QY 473 NLSHVTENLIGAFVPEQIIQERPL--TNKLRKRRPTSGHAPDETAKKADLAVQKPE 530
Db 878 DLGSSVE-----LOBEDLSHTQGSLSKNSPMAVEDVDAAHAAVOLTNRSD 924
QY 531 MINQGTQEQNGQVNNITNSGHNKT-KGDSIQNEKNPNPIESLEKESAFKTAEPIS 589
Db 925 SLAVVHDYTEKSRNNIEHQKGTEDKDFKSNLSLMMKSDGNSDCSDKMSSEF---LDPVLN 981
QY 590 -----SISNMELELNHNSKAPK--NLRKRSR--HIHMLELVSNLSP 634
Db 982 HNGGSFRTASKEIKELSHENVKSKMFKEDIEQYPRRLACIDIVNLPPLANOKISEP 1041
QY 635 NCTELOIDSCSSSEETKKKYNQMPVHRHSRNQLMEGEPATGAKKSNKPNEQTSKRHDS 694
Db 1042 HIEFLK-----SVTYVTSQSHQSSVSH-----EDT 1067
QY 695 DTFPELKLITNAPGSEFTKCSNTSELKEFVNPSPLPREKEEKLITYK-----VSNNADP 747
Db 1068 DTAPQM-ISSKODFHSNNLITQKAEITELSTILESGSQEFQFRKPSHIAONTSEVP 1126
QY 748 KOLMLSGEVLOT-ERSVSSSISLVPGDVGTQGISLLEVSTLGKAKKTEPNKCVSQA 806
Db 1127 GNQWV---VLSASKWKMDTDLPLVPDSVQDTHSKQFESA-GVAKSPHILLEDTGN 1181
QY 807 AEENPKGLIHGSKQNRNDTE--GFKYPLGHEVHNSRPTSIEMEESELDAQYLFNTFKVS 864
Db 1182 --KNT-----SCFLRNINEMERGFCALGTRKLSVNE-----A 1213
QY 865 KRQSFALFSPNGNAEECA-----TFSHSGS-----LKKQSPKVTFF--CEQK 906

```

```

Db 1214 LRKAMLFSDIENSEEPPSAKVGPRGFSASHDSVAVFPIKIKONTKESFDEKSSKQVY 1273
OY EENO-----GKN-ESNIXPVQ-----TVNTAGEPVV---GQK 935
Db 1274 LQNNIEMTTCIFVGRNPEKTIKNTKHEDSTSSQORNNLENSDQSMSTSG-PVYIHKGD5 1332
OY 936 DKPVD-NAKCSIKGSRFCLSSQFRNGELITPNKHGLQNPYRIPPLPIKSFVYKTC 994
Db 1333 DLPAQDQSKC-----PESC--TQYARENTQIKENISDLCLFELMAEETCMKSSDKQL 1385
OY 995 KMLLEENFEHSMSPREMGNGENISTYSTIRNNIRENVFEASSNINEGSSINNEY 1054
Db 1386 PSQKMONIKEFNIS-----FOTASGKMYR--VSKESLKNVNIINRREDE- 1429
OY 1055 GSSINEIGSSDENIQAELGRNRPKINAMRLGVLOPEVYKQSLPGSNCKHPEIK-KOEY 1113
Db 1430 ---LVIYISDS-----LNSKLHGIINKDKMH-----TSCKKRKAISTIKVF 1465
OY 1114 EEVYQTVNTDFSP-----YLISDNEQPMGSSHAS-----QVCSETPD---DLDDGE- 1158
Db 1466 EDHFPIVTVSQLPQAQHPXEYIESTKEPTLLSFHTASGKKVKIMQSLDKVKMLPDETQY 1525
OY 1159 IKEDTFAEED--IKSSSAVFSKSVOKGELSRSPPTHTHLAQYRRGAKKLESSENL 1216
Db 1526 VRKTAFSQSGSKPLKDKSKELTLAYEKIEVY-----ASKCEMQNFV 1567
OY 1217 SSEDDELPCQHLHFGVNNIPQSTRHSVATE-----CLSKNN---TEENL 1260
Db 1568 SKETEMLPQONHMYKOTELKTSNGTSKVOENIENNEKNPRICCIQSYPVVEDSA 1627
OY 1261 LS--LKNLSDCSNQVILAK-----ASQENH- 1284
Db 1628 LAYTDSRRTCYRESLSGKRWLRQDQKLGTRNTIKTECVKEHTEDEAGANASYEHSL 1687
OY 1285 -----LSEETCASLFSQSCSELDL----- 1306
Db 1688 VIIRTEIDTNHVSENVSTLLSDPNVCHSYLSOSSFCHCDMHNDGFLKKNKIDSDVP 1747
OY 1307 -----TAN-----TNQ-----DPLIGSK- 1322
Db 1748 DMKNAQNTISPRVSAATKERNLHPOTINECVOKLETNISPANKDVAIDPSLSDSNCK 1807
OY 1323 -----QMRHQSOGVGLSKDELVSDEBERGTGLEENNOEQSMDSNLGEASGESET 1376
Db 1808 VGLSIVETAHSQETE---RTKEIVTDNCKYIV-----EQNRQSKPTQCTQCHKVL 1855
OY 1377 SVSEDCSGSSQDILITQOORT-----MOHNLIKLOQEMAELEAVLEQHSQSPNS 1428
Db 1856 DSKRDIICPSSSGDVICINSRKDSCFPHNEQILOHN-----QSM5-----GLKKAAT 1901
OY 1429 YPSIISDSALEDLRNPED-----STSEKAVLTSQKSESEYPISONPEGISADK 1476
Db 1902 PRVGLTWTQTSKIRPRPOAHPSTRYGTJSTASGALIOVSDASLE-KARQVFSEMDGA 1960
OY 1477 FEVSADSTSKNKP--GVERSSPSKCPSLDRMYHSCSGSLQ-----NNYP----- 1523
Db 1961 KQLSMWSLSEGNKPRHVSXKRENS-----VYHSTQGVLSLKPPLGNVNSSVFSG 2010
OY 1524 -----SOEELIKYVDV-----EEQOLEEG- PHDLTETSYLPRODLESTPY 1563
Db 2011 FSTAGKRLVYSESALHKYGMLEFDLIRTEHTLHSPREDVYSILPQCAEIRTPREY 2070
OY 1564 -LESGLISLPSDDESPSEDARAPASRGVNIIPSTSAKLVPOLKVAESASPAARTTDT 1622
Db 2071 PVNSKLOKTYTNDKSSLP5NFK--ESGSSGNTQSIEVSLQSLQMERODTOLVIGTVSHS 2128
OY 1623 AG-----YNAEESVAREKPEL-TASTERYNKRMSWVSGITPEEFMLVYKFAKHHITL 1676
Db 2129 KAMILGHEQTLPNIKVKTDEMTKFDVPIKTVNGEYYSKESENYFETAVEAKAFMED 2188
OY 1677 TNLITETTHVWKTDAEPVCERTLKFLCIAGAKVWVSFWVYQSIKKEKMLNEHD--F 1734
Db 2189 DELTDEQTHAKCSL---FTCPQNETLFLNSRTRKRGVTVDAVGOPPIKSLNEFDRII 2245

```

```

OY 1735 EVRGDVVNGRHHOGPKRARES---QDRKIF---RGEICCYGPF 1772
Db 2246 ESKGSLT-----PSKSTPDGYTKDLSLFTHHMSLEPVTCGP 2283

RESULT 8
T14156
kinesin-related protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14156
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
Cell 91, 357-366, 1997
A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chrom
A:Reference number: Z17893; MIMD:98028574; PMID:9363944
A:Accession: T14156
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2954 <MOO>
A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1
C:Genetics:
A:Gene: XCBP-E
C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 3.7%; Score 355; DB 2; Length 2954;
Best Local Similarity 19.3%; Pred. No. 1.3e-07;
Matches 358; Conservative 276; Mismatches 678; Indels 546; Gaps 87;

OY 65 KNDITKSLQESTRFSLQVELLKIICAFOLDTG-LEVANSYNFAKKE--NNSPEH----- 117
Db 552 KEKIT--SLQOQLOSKR--EKKELVQSFELKIALEEQSLVAKNLEMYTNSREHINA 607
OY 118 -----LKQEVSTIQSMGYRNRKAKRLQSEPNPSLQETSL5----- 153
Db 608 EVQTDVEKEVVRKREMSVYLGSGY-----NASMSDLODSSVDGKRLSSHDECIEH 657
OY 154 -----VOLSNLGVTRTLRTKQRIQPKTSYVIELGSSSEPTVKATYCSGQDEL 205
Db 658 RKLAEQIVLLEF--TENLKKSENDKQ5-----SEDDFM5IQLC-----EAI 701
OY 206 QITPQGTREISL--DSAKKACEFSET---DYNTNEHNP5NDLTTERRAERPERK 260
Db 702 MARKANALELALMRDNDFNILE--NETLKEIADERSLKENQETNEFEILEKETQKE- 759
OY 261 YQSSSVSNLHVEPCGITHASSLQIHENSLLIKDRNNVKAEPCKNSKQPG-----LA 314
Db 760 -----HEAQLIHETIGSL-----KKLVENAEVYNQNLDEDELETKTKL 796
OY 315 RSQHNWAGSKETCNDRTRETEKKV---DL-----NADPLCEKKEWKKOLPCSEN-PR 365
Db 797 KEQEIQLAEIR-----KRAVNLQKVVNFDLSVMSGSEKICEITPOLKQSLSDAEAVTR 851
OY 366 DTED-----VPVITLSSIQKYNWFSRSDDELGSDSHDGESESNK-----VAVL 413
Db 852 DAQKECFELSENLELKEKEDTSMYNOKEKAASLFEKQLETKESYKKAEMADLOQELQ 911
OY 414 DVNLNEDVYSGSSEKIDLLASDPHEALICKSE-----RVHSKVSYESNIEDEKIGTKYRK 468
Db 912 SAFNEINLYNG--LLAGKVPRLDSRVELEKKVSEFSKOLEKALEEK-----NALENE 962
OY 469 AS-----LPN-----LSHVTEMLIIGAFVTEQIIQIQRPLTKLKRKRRTSGHL 513
Db 963 VTLCEKFLPLNVEYECIKNOISKA5EITL-----LQOE-----GEH 999
OY 514 PEDFIKADLAVQKTEMINQGTNO--TEONGOVANITNSCHENKTKGDSIQNEKNPNPIE 572
Db 1000 SASITISQEIIMQESQIQLITDEVYHTQSKVOQTEGYLEMKKMHDL----- 1049
OY 573 SLEKESAFKTAEPRISSISNMF-----LELNTHNSKAPKKNLRRKSSRTHALELVY 627
Db 1050 -PEKYIRNKSSEADLLKREMNELKGTME5VEVKIADTKHELEETIRKEQOLH----- 1100

```

```

QY 628 SNNLSPNCTELQIDSCSSSEIEIKKKRYNOMPYRHSRLQIMEGKEPATGAKKSNKPNQ 697
Db 1101 -----EKKY-----FQAMOTIPPIPL----- 1118
QY 668 TSKRHSDFPELKLITNAGSFTKCSNTSELKEFVNPSPREKKEKLEVKYKSNNAEDP 747
Db 1119 -----SDLPPSKLVEG-----NSODPIEINDYHMLIALATERNNIMCLETERRNSLKE 1167
QY 748 KDLMLSGEVLQTERVESSSISLVPGTYGTQESILLEVSTLGAKEPTKNCVSCAA 807
Db 1168 QVIDANTQLOAQISQIEKSDQ--KKQDLEBEVALLME----- 1208
QY 808 FENPKGLHGCSKDNKNDTEGFRPLGH--EVNHSRET--STEMESELDAQYLQNTKRVSK 865
Db 1209 -----LTKGHILDSOLIE--KLQENLEVEKTLQDQEMKNITIERNELQNTEDLK 1260
QY 866 ROSFALFSPGNAEECAFFSAHSGSLKQSPKVTYECQKEENQCKNESNIPVOTVNI 925
Db 1261 -----AEHDSLKOD--LSENIQSIEETQDE-----L 1284
QY 926 TAGFPVQCKDPVDNAK-----CSIKGGSRCFLSQFNGNETGLTPNKHGLQMPYR 980
Db 1285 RAAQBELRQKQOLVVSFRQQLDCSV-----GISEPN--HDAVANOQEV 1326
QY 981 PPLPFIKSF-----YTKCKKNLLE--ENFEHSMSPERKNGENIPSTVSTISR 1028
Db 1327 --SLGEVNSIQSEMLRGERDELQTSCKALVSELELLRAHYKVSQGE--NLEITKRLNGLEK 1383
QY 1029 NNIRENVFEKASSNNINEVSSSTNEVGSISINEGSSDENIQAEILGNRQPKLNAMLRLGY 1088
Db 1384 ELTGSESESEVYKSLMELNKEKNKKEAEVSSKQENQPSLEEVSGSKL-----Y 1436
QY 1089 LQPEVYKQSLPGSNCKHPEIKKQEEVYQVNTDFSPYLLISDNLQPMGSSHASQVCE 1148
Db 1437 DEIEVLKAKOLKAAE--ERLEIKRDYFELVOTANTN--LVBEKLETPLOADHE--ED 1487
QY 1149 TPDDLQDGEIK-----EDTSPAENDIKESSAVFSKQKQELSPSPFTTHLAQGY 1202
Db 1488 SIDRSESEIEIKVLEKLEKERNQYLLERLOEKELELSNKLEILOKEMETSVLLKDDQO-- 1545
QY 1203 RRGAKKLES-----SEENLSEDEDELPCFOHLLFGKVNINIPSSQ 1241
Db 1546 -----KLESLSLENILKENIDTTLKHSHTQAOLOQTOEOLOLAKNLIAASDNCP--I 1598
QY 1242 TRHSTVATECSKNTENMLSLKNSLDCSN--QVILAKAQEHHLSEETKCSA-----S 1294
Db 1599 TOEKETSDACVHP--LEEKILLTLELHQKTNEQKELHKEKNELEQAOVLEKCEVEHLKMS 1657
QY 1295 LFSSQCSLEDLTANTNODPPLIGSSKQMR-----HQSQGY 1333
Db 1658 MIESK--SLESLOHEKHDTEDQLLAKQOQVVOYTEKKELQOTHEHLTAEDHLKENIEL 1716
QY 1334 GLSDKELVSDDERGTG-----LEENNOEPOS-----MDSNLGEAASGESSTVS 1379
Db 1717 GLNFR-----NEAQKTKTEQCLLENKELQESQHRLOCEIEBLKSLKDESALETIKESB 1773
QY 1380 EDCSGLSSQSDILITQ-----QDRTMOHNL-----IKIQEAMAELEAVLE 1419
Db 1774 QKVINLQEMEMVLEMELEKNSQRTVIAERDQLODQDLRESVEMSIETQDQDLRAQEAQ 1833
QY 1420 OHGS--QPSNVPSTIISDSAL-----EDLRNPEST--SEKAVLT-- 1456
Db 1834 QOKDKVQELTQSOISVLOEKISILENOMLYNATVKEITLSEERDQLOSKOHLSEIETTL 1893
QY 1457 SOKSESEPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPKCSLDDRY--MMS 1512
Db 1894 SLKEFEALQO-----AEKDKADAKKTIDITEKISNIEBOLLQOATNLKETLYERSLIQ 1949
QY 1513 CSGSLQNNRNPQOELLIVYVDE-----EQOLESGPH--DLTE--TSYLRQDLEGPYLE 1565
Db 1950 CKEQDLA--LNTHELRLTAKSKDALGKMQEOREAANKYALVLEKMSLSLEQINENTVTLK 2008
QY 1566 SG-----ISLFSDDPESDPESEDRAPESARVGNIPSSSTALKVPQLKVAESAQSPAHAHTTD 1621

```

```

Db 2009 ECEGEKETFYLPQPSKQSSQME-----LRESIKTKDLOL--EEAKEISEATNE 2058
QY 1622 ----TAGYNAMESSVREKREBELT--ASTERVKNRMS--MYVSGLTPEEPMLYKKFARKH 1672
Db 2059 IKMLTAKISLEEELIOMNNSILNEAVSERENLRHSQOOLVSEL--EQLSLTLK--SRDH 2113

RESULT 9
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohba, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratazu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BAE43752.1; GSPDB:GN00149
A:Experimental source: strain N315
A:Genetics:
A:Gene: SA2447

Query Match 3.5%; Score 342; DB 2; Length 2271;
Best Local Similarity 18.5%; Pred. No. 3.3e-07;
Matches 310; Conservative 266; Mismatches 835; Indels 266; Gaps 49;

QY 161 TYTTLRTKRIOPKTSVYIELGSDSEEDTVKATYCSGDOELQITPQGRDEISDS 220
Db 638 TYTVVSTQDANKKSTTTITIN--VDTAPTIV-----PIGQSSSEVSP-----ISP 683
QY 221 AKKACESETDYVTEHHQPS-----NNDLNTTEKRAERHPEKYQSSVSNLHVEP 273
Db 684 IKIATQDNGNAVTVMTVGLTFGLFDSNTNITISGT-----PTNIGTSTISIVSTDA 735
QY 274 CGTNTHASSLOHENSLLTKRNVKAEKFNCKKQKQGLARSOHNRKAGKEKCNQ--- 330
Db 736 SGNKT--TYTFKYE-----VTRNSMDSVSTSGSTQOSQSVSTSKADQSASTSTSGSIV 789
QY 331 RRPSTTEKKVVDN--ADPICKERKMKOKLPCEPNRQEDVPMWTLNLSIO--KVNEMFSR 388
Db 790 STSASTSKSTVSLSDSVASAKSLSTSESNSVSSSTSLVNSQSVSSMSGVSKSTSL 849
QY 389 SDELLGSDSDHDESESNKAVADVLDVLEVDE-----YSGSEKIDLLADPHEALICK 443
Db 850 SDISINSNSTERKESLST--TSDSLRTSTSLSDSLSMSTSGSLSKQSLSTIS-- 902
QY 444 SERVHSKSVESNIEDKIRKTKKASLPNLSHVENLITGAFYEPQIIDERPLTKLK 503
Db 903 ----GSSSTVSALST--SNAISTSTSLSESASTSDSISTNSIANSQSAS--TSKSD 953
QY 504 RRRRPTGLHPDPIFKADLAQVQKPEMINOCTNOTEQONGOVMMNTNGHKNKTKQSDIO 563
Db 954 SOSSTIS-----LSTSDSKMSMTSESLSDSTSGSVSGSLAIASQSVSTSTSDMS 1006
QY 564 MEKNPNPIESLEKSAFKYTAEPISSTISNNELELINHSKAPKKNLRKRSSTRHIIA 622
Db 1007 TSEIVSDSISTSGSLASDSKMSVSSMST-----SQSGSTSESLSDQSTSDSDS 1058
QY 623 LELVYSRNISPNCLELOI-----DCSSSEELIKKKRYNOMPYRHSRLQIMEGKEPA 675
Db 1059 KSLSLSTSGSTSTSTSTASVRTSESQSTSGSASQSDSMSTSTSFSDSTSDSKSAS 1118
QY 676 TGAKKSNKPNQETSKRHSDPPELKLITNAGSFTKCSNTSELKEFVNPSPREKKEKL 735
Db 1119 TASSBSISQASASTSGSVSTSTSLSTNSERTSTSVSDSTSLSTSESDSI--SESTST 1176

```



```

QY 736 ETWVKNNAEDPKDMLSGERVLOTERVSS---SISLVPGTGYQTOESISL----- 786
Db 1177 DSIISAIKASESTISLSESNSTSDSESQASAFISELSESTSESTSESSVSSSEST 1236
QY 787 ---EVSITLKATEPKNKVCOCAPENPKGLIHGSKDNKNDTEGFKYILGHEVNHKRT 843
Db 1237 LBDSTSESGSTISLSTNSGASISTSTISSEST--PFKSESVSTLSLMSSTISLST 1295
QY 844 SIEMEESLDAQYLONTFKVSKROSFALFSPNGMNEECATSAHSGSLKKQSPVTFPC 903
Db 1296 SILSTLSBDSKSDSLSTMSSTSDIST---SKSDSLSTSTSLSGSTSES-----ES 1346
QY 904 ECKEENOGKESNIRPVQTVNITAGPPVYGKDPVNAKCSI-----KGSRECLSS 956
Db 1347 DETSSESKSDSTMSISMSOSTSGTSTSTSLSDSTSTSLSLASMNQGVSNMS 1406
QY 957 QPRGNETGLITPNKGLONPRIPPLPKSFV-----KTCKKNLEENFEHSHS 1009
Db 1407 QASASTSTSTSESDSOSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1466
QY 1010 ---PEREMGENIPSTVSTISNNIRENVFKASSNINFGVST----- 1051
Db 1467 LSGSESESSOSTISASSESTSEASTSLDSTSTNSGASTSTSLNSASASESDSS 1526
QY 1052 ---NEVGSSINEIGSSDENIOAELGRNGPKL--NAMLRLGVLQPEVYKOSLPGSNCKHP 1106
Db 1527 TSLDSTSTASMSQSSDSOSTSASLSDSLSTSTSNRMTIASLSTSV--STSGSTSEST 1585
QY 1107 ELKKOFEVYQTVNTDPSPL-----ISDNLEQPMGSHASQVCS 1147
Db 1586 SESDSTSTSLSDSOSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1645
QY 1148 ERPDDLDDGELEKEDTSFAEDNKESSAVFSKYOGKELSPSPPTHTL----- 1198
Db 1646 STSTMSDSTSLSDVSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1702
QY 1199 -AOGYRGAKKLESSEENLSED-----EELPCFOHLECK 1233
Db 1703 DQSSSESVNDESESVSESNESDSKMSGSTSVSDGSLVSTSLRKSSVSESSLGS 1762
QY 1234 VANINOSQTRHSTVATECLSKTEENLSLKNLSDCSNOVLLAASQHHLSEETKCA 1293
Db 1763 QMSDVSSTSDSSLSVSTSLRSSBS--VESDLSLDSKSTSGSTSTSTSTSTSTSTSG 1821
QY 1294 SLFSSQCELEED---LTANTNTQDP-----FLIGSKOMRHOSEQVGLDKEIVSD 1343
Db 1822 SESVSESTSLSDSISMSDSTSTSDSLSGSTSLSTSDLSDSKSLSSQSSMSG 1881
QY 1344 DEERGTLGEENNOEBS---MDSNLGEAAGCESESTSVSEDCSGLSOSDILTTQORDM 1400
Db 1882 SESSTSTSVSDQSSSTSNQPFMSISASESDSMCTSDOSNISGNSSTSTSLST--DSM 1939
QY 1401 QHNLIKLOQEMAELEAVLEOHGSOPTNSPTIISDSSALEDLRNPDESTSEKAVLTOKS 1460
Db 1940 SGSV-----SVSTSTSLSDSISGTSVSDSSSTSTSTSLSDMSOSOSTSTASGSLST 1994
QY 1461 SEYPTSONPEGLADKFEVSADSTS-----KKEGCVRRSSPKCPSPIDRMV 1509
Db 1995 ISTSMMSASTSSSOSTSVSTSLSTSDSTSLSTISISGOSTVESESTSDSTSDSES 2054
QY 1510 MH--SCSGSLQNRNYPQOELLKVVDVEQOOLEESGPHDLTERTSYLPRODLESTPYLESQ 1568
Db 2055 LSTSDSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 2107
QY 1569 SLFSDPDSPEDAPESARVGNTPSTSAKLVQOLKVAESAQPAANHITDTGYNAM 1628
Db 2108 SMSQKSDSTISDE-----SVSTSTST-----SLSTSDST--STSESLSTSMGSSOSI 2155
QY 1629 EESVSRERKELTASTERVNRKMSVYVGLTPEEFMUYVYFARKHITLITLITEET 1685
Db 2156 SDSTSTSMGSTSTSES-----NSMHPSDSMH---HTHSTSTSRSLSEATT 2200

```

```

RESULT 10
T41023
probable nuclear pore complex-associated protein - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 03-Dec-1999
C:Accession: T41023
R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Accession: T41023
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1837 <MUR>
A:Cross-references: EMBL:AL023860; PIDN:CAN19588.1; GSPDB:GN00068; SPDB:SPCC162.08C
C:Genetics:
A:Gene: SPDB:SPCC162.08C
A:Map position: 3

```

```

Query Match          3.5%; Score 341.5; DB 2; Length 1837;
Best local similarity 19.4%; Pred. No. 2.6e-07;
Matches 351; Conservative 310; Mismatches 674; Indels 475; Gaps 78;

```

```

QY 2 DLSALRVEEVQNVINAMOKILECPI-CLELIREPVSTKCDHIFCKFMKLTKORKQPSQ 60
Db 130 DQNLKQTAQLQNLNLSDKKEVEKKTITIKDLALASSHQV-----LELQHTQO---- 179
QY 61 CPLKNDITKRSLQESTRPSQVLEBLKTKAFOLDTGLEVANSYNFAKE--NNSPEHL 118
Db 180 -----EKASLQNTYEP-----ELQKLT---QKNSILNNTWLSRELQGVNDKLSTL 223
QY 119 KDEVSIQSMGYRNRKAKRLQSEPNPSLOETSLVSQVNLG---TVPTLR-----T 167
Db 224 HOEASILEKQ-----LSSQLDAVLEKALQKRYSLSQOTTESLRLPQNIYALIS 274
QY 168 KQRIQPKTSVYIELGSDS---SEDTYRKATYCSV-----GDOELQITPGTRD 214
Db 275 EMRKQYEFQVAFKEIISQKQISELMMEKCDCLRLKEIQNSGELKLELAQSSPE 334
QY 215 ELISLDSAK-----KAACFSETDVTNTEHHPNSNNDLTTTKRAERPEKYQSSSVN 268
Db 335 E-QLESHEKAEASLSQINFLEKEVSSLE-----SOLKANERL--RHYDEIETSDME 385
QY 269 LHVPCGNT-----HASSLOHENSLLTTRD--MNVEKAFCKNS-----KOPG 312
Db 386 LKYSNLNLSMKGFGQSSVDLYSERLYKKQTEQCEVRLORSTYHWEANLOHP 445
QY 313 LARSOHNRWAGSKE---TCNDRRTPESTKRYDLNADPLCEKREKMKOKL-----PCSEN 363
Db 446 LKEDQFKRAHQREIVAMSEQYQKSLD-----QKAKSYEQLETLFKKKTJEN 496
QY 364 ---PROTEDVP-----WITLN--SSIQKVNEMFSRSDDELGSDDSGESE 404
Db 497 KHVEQETKDLARQOVULLHELDICENGIVLGVDSKRTKINSYEKS---LTBEDTDTDI 553
QY 405 SNAKYA-----DVLVDLVNDEYVSSSSKIDILASDPREALICKS----- 444
Db 554 SRLVYFNRIRLQOONOLSAVHELADRMKEDKPDLDGAEIOEETLIKANETIDOLT 613
QY 445 ---ERVHKSYSVNSNEDIKFGTYRKKASLPMLSVHTENLIGAVTEP-----OIIOER 496
Db 614 KMLEVSDQLKRYSLKERDPFRSLVQENKELDMAPATPNSKNTNLIDQTSYQRLIRE 673
QY 497 PLTNLKKRRKRTSLGHEPDIKKADDLAVOKTPEMINOGTNOTEONGOVANITNSGHENK 556
Db 674 QLTNLELSKTSIRNKE-----KKEFEAISL-----OLEKSNLOLOLTSLTSR- 718
QY 557 TKGDSIQNEKNPNPESL-----EKESAKTKAEPIS--SISNMELELNTIHNSKA 605
Db 719 ---SLAEKLNDELKSLVLSRSDDELDESKLSQEOBASKKIEVQVYSSOLSICNSOL 774
QY 606 PKKNRL--RRKSTRHIALLELVSRNLS--PPNCTELOIDSCSSSEIKKKKYQNMQVR 661

```

```

Db 775 EQSHIYDNLKSENLLTTSYKOKLKLADLSNESKLSLQOQNFHMAKQIESSNOEYATV 834
QY 662 HSRNLQMEGKEPATGAKKSNKPNQOTSKRHD-SDTPPELKLJTNAGSF-----T 710
Db 835 DSMNSRIE-----LSNDLRVANSKLSLSCSDVRLTIQN---SPDLREHOTLVQ 882
QY 711 KCSNTSELKEVNPDLPR-----EKEEKLKYKVSNNADPKDL 750
Db 883 LQSNITELKODI--TLQRTVRNOLEIQTTELKERLKFMEERENQSKITIAN---KDT 936
QY 751 MLSEERVQTERSVSSISLVPGDYGTOGSISSILEVSTIGKATEPKNKCVSOCAFEN 810
Db 937 TQNDNV-----EVAISIEL-----ERTKEKLMAE--LEKSNIOQYKLASE----- 977
QY 811 PKGLIHGCKDNNDT-EGFYRPLGHEVNSRE-TSIE-----MESEIDA 854
Db 978 -KTL-----EMMNETHEQFHLVSEISTREKITSLSRELLDKNKREVLKEKES 1030
QY 855 QYLONTKRVSKRQSFALFSPNGNAEECATSAHSGLKQSPK--VFEECEQKEENO- 910
Db 1031 KELKQLEDAVREKDSALSFKKDKYKIRSDADRVITSKEDIKERSLMECHSNVESEI 1090
QY 911 ---GKNESNIPYOT-----VNITAGFPV-----GQKD-----KPYDNKCKSK 947
Db 1091 VSHERTQKLDLRTEDDEVNTKYLLKLANFEQOHSGLSGAEKDNINORAKMEQETISLK 1150
QY 948 G-----GSRFLCSOF-----RGNETGLITPNKGLLQNPYRIPLPIKSF 989
Db 1151 DYILGLENQNKLLHSDPLSLSQOITVLOQNSSENNITANLEAVQDNDR-----ELVSY 1205
QY 990 VKTKCKANLEENEEHSMSPREKMGENTIPSTVSTIRNNIRENVEKEASSN----- 1043
Db 1206 LRH--EKEIMDNKY-ELTILDNRGL-NOOVKSLQSTVSLQLELRLSLPLSPNSQOTDP 1261
QY 1044 -----INEVSGSTNVGSSINEIGSSDENIOQELRN 1075
Db 1262 IISGQEVQLLYESNSVLRKNDQKLGKIQELEKEVEKLANSLQLEITINELKKEIQAK 1321
QY 1076 RG-----PKLMLRLGLVLOPEVYKQSLPGSNCKHEIKKOEYEEVQ 1118
Db 1322 TASLNLMEKYNRMKLRFOVLYNKERYVDPQLEELK-----NCEALEKEKQOLEETLQ 1376
QY 1119 TVNIDESPYLISDLQPMGSHASQVSETPDOLLDDGELKEDTSPFENDIKESSAVFS 1178
Db 1377 ETAKE-----TDFKQOVNS-----LNEEVENL-----KKEVEQANTKTRLAAMN 1418
QY 1179 ---KSVQGEISRSPPTHTHLAOGYRGAKKLES---SEENLSEDEBELPCFOHLFG 1232
Db 1419 EKCENLKSSLTR-----FAHLKQELTNKKKELTSKAENEMOKETIESLKDSSHQLOE 1472
QY 1233 KVNNIPQOSTRHSIVATECLSKNTEENLLSKNSLNDSCSNQVITAKASQE-----HH 1284
Db 1473 SASDAEQITKEQFEQKSEKEREKELADSKNELEHQSEAVDADGTEISNLEKEIHE 1532
QY 1285 LSEETKCSASLFSQCSFLEDLTANTNTQDP-----LIGSSKQMRHOSSESGV 1333
Db 1533 LRSQKEGLVOOVMLSAELALRHSPTQSGLENADETARLSQLESTKYOEKKEKETEI 1592
QY 1334 GLSDEKLVSDEEGTGLEENNOEQSDSNLGEAASCESETSVSEDCSGLSQSDSLT 1393
Db 1593 LAASELVAEKEKKELEENLNEKSQRIKELBQAQNSSENT-----HNDIDDTIK 1645
QY 1394 TQQRDTMOHNL---IKLOEMALE-----AVLEQHGSOPSNSPYIISDSSAL-ED 1441
Db 1646 QOYVEKELKENSANDVLLKVVAFETEPRSKAKISYEKKTDLQNKITQLEETIENLKKQ 1705
QY 1442 LRNPQOS-----TSEKAVLTSOKSSFTPISONPEGISADKFEVYADSTSK---NKE 1490
Db 1706 LSNPEKDESTSVTETKPYTSKPTASKADVQONATEASSAKRPSGKSLARLQGTGKQ 1765
QY 1491 PGEVRSPPK 1500
Db 1766 KGVORPAVSR 1775

```

```

RESULT 11
JC5837
364k Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
C:Tokl, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misum, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein
A:Reference number: JC5837; MUID:96093490; PMID:9431462
A:Accession: JC5837
A:Residues: 1-3187 <PRO>
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Cross-references: DDBJ:D25543; NID:9516825; PID:BA05026.1; PID:9516826
C:Comment: This protein plays a role in the formation and maintenance of the characte
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status pred
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

```

```

Query Match 3 5% Score 340.5; DB: 2; Length 3187;
Best Local Similarity 18.9% Pred. No. 6.2e-07;
Matches 375; Conservative 326; Mismatches 719; Indels 561; Gaps 91;

```

```

QY 23 ECPLELTKPVPSTKCDHIFCKCMYKLLNQKKGPSQCPICKNDIKRSLOSTSPSL 82
Db 919 EEPVKEALQOLE-----WLRKRSQRRKRLQAL---ISKLELO--KYSKL 962
QY 83 VEELKITICAFOLDGLEVANSYNFAKKNNSPHLKDVEAIIOSMGYRRARRLQSEP 142
Db 963 EEELAKV-----RESETKDSL-----RESEKRELEDS 990
QY 143 ENP-----SLQETSLSVQSLNLTGVTRTKRKQRIQOKSVYIELGSDS 186
Db 991 KKKDDPEKVTGSEMELEFVSLRLTISEKEVELEIRDLKKAABEELQALVQRM--- 1046
QY 187 SEDTVNKAATYCSVQDOLL--QITPQ-----GTREISLDSAKKACFSETDTNTEHHQ 240
Db 1047 TQDLQNKTKQDLDLQEELETENQATIQKFTGTMTAGGDSAVKETSVPSPRAGGEGHMK 1106
QY 241 PSNNDLNTTEKRAARHDEKTYGSSVSNLHVEPCGTHTASSLOHENSILLTKDRNNE 300
Db 1107 PE-----LECKIVDLKEKTYO-----LOKKQLEALLISRKAI-LK 1139
QY 301 KAERCNK-----SKQPLARSQHNRAAGSKETCNDRRPTSTEEKVQDLMADPLCEKRENN 354
Db 1140 KAOEKEKHLKEELKEQKDADRHLQDFQSGKNEENIRAPLRO-----LOAKEST 1189
QY 355 KQKLPCSENPREDTVDVWITLNSSTQKVNEMFWSRDEILGSDSDHGES---ESNAKVAD 411
Db 1190 DQQLPGTQOQPHPTGSELSLEGT-----EPASSD--LHMAQPSHPGETITLQATYSVQ 1243
QY 412 VLVDLVNEYDEYSGSSEKIDILASDPHEALICKSERV----- 447
Db 1244 IQDOLKEIEV---EKEELEKISSSTSELTKKSEVVLQLEQIINEQGLEIQLNKAASHBA 1300
QY 448 --HKSQVSNIEDIIFPKTYRKASLNL-----SHYTE-----NLIGAFVIE 489
Db 1301 KAHTEQKQOLE-----SSQIKTADLEHLTLQLELETQKHGQKEEVSYLVGOLGER 1355
QY 490 PQILOERPLTKLKRKRPTSGLHP-----EDFIKKRADLV---OKTPMINQGTN 537
Db 1356 EQTL--TTVQTEMEQERLTKALHTQLEMAKKEHEKLVQVYELCELKQKPELEDESK 1413
QY 538 QTEONGOVMMITNSGHNKTKGDSIQNEKNPNPTLESKE-SAFKTKAEPISISSIMEL 596
Db 1414 AKQO-----LQKIQALALISRKALKKESKSLQLOESSARDAVEHLTKSLADVS 1462
QY 597 ELNTHNSKAPKKNLKRKSSRTHLHLELVYSKLS-----PPMCTELQIDSCSSSE 649
Db 1463 QVSVOQO---EKDALLGKLLALLOERDKLIVEMDKSLLENOSLGSCESLKLALGLGTED 1519

```



```

QY 425 SSKIDLL-----ASDPEALICKSERVSHSKSVESNIEDKIFGKTYRRKASLPNLSHYT 478
D 1065 IKRDVDELNVNQVISENKVTLTFKNNSVYIEAMSHINTVAHGITSNKNELIKSKEVE 1124
QY 479 ENLIIGAVTEPQIIQERPLTKLRKRPTSGLHPEDFIKKADLAVOCTPEMINOCTNO 538
D 1125 DKL-----NLVEONEDYKKVKNPEN-----EKOLEAIRGSMSTLKEVINHVSE 1168
QY 539 TEONQVANNITNSG-----HE-----NKTGCD-----SIOEKNPNPRESLE--K 576
D 1169 MTQLESTANTLKSNAKGENEHLDELNKTQOMRDITYEKLLIAELKEGYNELKADN 1228
QY 577 ESAFTKAPISISSISNMLELNIHNSKAP-----KXRLRR 613
D 1229 EKANVPEPERNIIIGHYLERITVEKDRAGKAYEEMNSLKTIEKLIQETSDOSQELVY 1288
QY 614 KSTPFIH--ALVELVSNLSPNCTELQIDSCSSEIK-----KKTKNQMPVNH 662
D 1289 TSTKHLNNAKQEVYIKRN-----EEDSIOLREKAKSLETLEDKKKLVQVY---- 1335
QY 663 SRNLOLMGEBEPATGAKKSNKPNE-----QTSKRDSDTFPELKTUNA 705
D 1336 --NMMLQAIQGNATISK--ELNELKGYIELLISTVYSSILEYKKNSESVRFSQLAN- 1390
QY 706 PGSEFTKCSNTSELKEFPNPSLPREEKEKLETIVKVSNNADPKDMLSGEVLQTERSYE 765
D 1391 -GEFKAEG-----BEKNASARLAERAKLEQIVKLDYSDIDDYKKIEGIRKELIKK 1444
QY 766 SSISLVGCTDGTQDEISLSEVSTLGRKATTEPNKCVSCAFENKGLIHGSKRNRD 825
D 1445 ESALTVEESEKFKOMCSSHMEKNAEGKKIETLK-----NN 1481
QY 826 TEGFYPLGHEVHNSRETSIEME-----SELDACIYQNTFYKSKROSFALFSPGNABE 880
D 1482 GDGGR-----ANITDSQMEVGYNTSKAHEHYEAQYDKTKAC-----E 1523
QY 881 ECATFSAHSGSLKKOS--PKVTFECEQKEENQKNESNIKPVQYNTAGFVVGQKDKP 938
D 1524 SIVAVYTKMDNLFNSELMEKVEYKCEKKDEAKYSAKLKP----- 1564
QY 939 VDNKCSIKGSRFLSSQFRNGENGLTPNKHGLLQNNYRPLPFIKPYKTKRCKNL 998
D 1565 -----YDGRIRAVSENERKI-----SELKEKAKVE- 1590
QY 999 LEENEHEHSMBERMGENENIPSTVTSIRNNIRENVFEKASSNINEVGSSSTNEGSSIT 1058
D 1591 -----KRESSQLNDVSTKSLQIDNCROOL--DSVLSNIGRKQ-----NML 1630
QY 1059 NEIGSSDENIQ-----AELGRNRG--PKLMAMLRLGVLOPEVYKQSLPSGNCNKHPEIKOE 1112
D 1631 QYFDSADSKMSVLPRISELGAESKSLDKVA-----AKES 1664
QY 1113 YEEVQVQNTDESPYLLIDNLEOPMGS--SHASOVCEPDDLL-----DDG---EIKED 1162
D 1665 YEKNETLEVQNEKSR1-----NWE--GSLTDIDKITTIDENDLLMKKQYEEGLQITKEN 1718
QY 1163 TSPFEND-----IKESSAVFSSVKO-----GELSRSPSP 1192
D 1719 ADKRKSNFELVGSSELNALLDPTSTIFIKLKEKYDMTGKLYGYKAMNIEHFEFTKSYN- 1777
QY 1193 FTHHLL-----AQYRGAKKLESSEENLSEDEE-----LPCFO 1227
D 1778 LIETFLSNATDYSVTFEKAQSLRELAEK---EEHILRREBEAIFLLNDIKKVESLKLK 1834
QY 1228 HLLGCKVNNIPQSQRHSTVATECLSKNTENLISKNLNCOS---NQVILAKASQCH 1283
D 1835 EMMKVSAYEYGMKRDHTSVSOLVODMKITIVDELTKLNDISECSSLNVNVAIVYKVKES 1894
QY 1284 HLESETKCSALFSSQCELEDTLANTNTQDPFLIGSSKOMRHQSPQ-----GV 1333
D 1895 KHADYRRAANSYEMSVT-----LANYFLSDAKIISGMEFNAEMKSNFKTDLLEITFSV 1949

```

```

QY 1334 GLSDKELVSDDEERGTELENNQOEQMSDNLGEAASCESETSVSEDCQGLSSQSDILIT 1393
D 1950 ISNSNELLKRTBQDSNDVYQKRESEQLAKDQATDIYVYIKLKNENEKLEFAKKNKEEYVS 2009
QY 1394 TQOROTMO-----NHLIKLOQMALEVLVDHGSQPS----- 1446
D 2010 EKVREALKRLSQVEGIRCFHEFNHRLDTELEMLKMKVNTYTRDKKSEERESQLOEMENE 2069
QY 1427 -NSYPS-----IISDSSALEDRNPEOST-----SEKAVLTSOKSSE-----Y 1463
D 2070 NMTYSNITQLEGIVYASAGESKEDIEKLERSNMEKRNISEKSTIDSIVIEMNSTIDELY 2129
QY 1464 PISQNPGLSADKFEVVSADSTSK-----NKEPVERSPSKC-----PSLDRMYVHS 1512
D 2130 KLGKNCQAMWISLISTYAMTKSKKLIMINK-----EKENTEKCVQYIKDNSSSTDGUYET 2185
QY 1513 CGSGLQNR--NYPQOEELIKVVDY-----EEQLESGRPHDLJETYLPROD-----L 1588
D 2186 LKGYFGSKLTFFSSASITVQNAPTYVNFAPKHEKESL--NAIRDLIKELYLPHQNSDISIV 2243
QY 1559 EGTPLYESGISLFPDDPESDEBPASAR--VGNIPSTSLAKVPQLVAESAQSPAA 1617
D 2244 EGG--VQNNMLALY-----DKLNEKREMDLYNINSETKLKQME----- 2280
QY 1618 HTTDIAG-----YNAMESVSRKEPELTASTERYKRN-----SNVYSL--TPEEMLY 1665
D 2281 HSTDYFKPMIELHKGMETNNKSLLEKEKKLSVNDHMSMEAEKMLNGLYTPEESYONI 2340
QY 1666 -----YKPAKRNH-----ITLITN-----LITEETTHV 1688
D 2341 NNITYIEAEVKTLEIDRDYDNTQIVEEHKQPSILIDRNALMDIEIFKKNENINL 2400
QY 1689 MKTDAEFVCERTLKYFLGIAGGKRV--VSYPWYQSIKERKMLNEHFEVGRGVNGRN 1745
D 2401 MEVNETTI--HRVNDYIEKT--NKLQVAKTEYQILLENIKONDMDQNIETKVSII--EY 2456
QY 1746 HOGPKRARES 1755
D 2457 FENVKKKRES 2466

RESULT 13
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Pulton, B.; Wohlmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: 221521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FNU>
A:Cross-References: EMBL:U080022; PDB:AA025885.1; GSPDB:GNO00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 3.5%; Score 339.5; DB 2; Length 3488;
Best local similarity 17.8%; Pred. No. 7.8e-07;
Matches 328; Conservative 302; Mismatches 733; Indels 475; Gaps 68;

QY 50 KILNOKGPGSQPLCKNDITKRSLOESTRFSQ--VEELIKITICAFQDGTGLEANSYNFA 108
D 367 KIIKKKEP-----ESQVYTAABEQOKISEVDQSVAFTEVGAKKKPKDAEKPTDLSKA 420
QY 109 KKEENSPEILKDEVSIIISMGYRNKAKRLLOSPEPNPSLOTSISVQSLNSGTYRTLRK 168
D 421 KDKSKSKSDPEASTE-----KSTTEKPTNDKTS-----KSAEKRTVYKPK 463

```





Query Match 3.5%; Score 334.5; DB 2; Length 3328;  
 Best Local Similarity 18.9%; Pred. No. 1.2e-06;  
 Matches 399; Conservative 286; Mismatches 724; Indels 701; Gaps 96;

QY 72 SLDSEFRESQVLELLKIIC-----AFOLDTGLEYANSTYPAKKNNS---PEHLKD 120  
 DB 466 SLDSEFRESQVLELLKIIC-----AFOLDTGLEYANSTYPAKKNNS---PEHLKD 120  
 QY 121 EVSIIOSMGYRNKAKRLLOSEPENPS---LOETSLVSOLNIGTYRTLRKRIQPK 176  
 DB 526 ACGLGILTRACQREDISICPSVDGSMPTTLDTSTAVK--NAGLSTLKNKR----- 577  
 QY 177 SVYIELGSDSEEDTVKATYCSVGDELQIQTQGTRELSIDSAKKADEFSETDV--- 233  
 DB 578 -----KFIY-SVSDASISQKKLQTHRQLELNL--SAQLENAFAFEVPL 619  
 QY 234 -TTEHHOPSNNDLNTTEKRAERHPEKYGOGSSVMLHVEPGCTNTHASLQHNSSL- 290  
 DB 620 FTWVNSGIPDSSD-----KKRCLEPNDE-----EPSLTNSFGTATSKELISYVHA 663  
 QY 291 LTRKDRMYEKAFCNKKSKOPGLARSOHNRWAGSKETC--NDRRTPETEKKVLDNADPCE 349  
 DB 664 LISQD-LNDKEALVITEKPOPTTARADPLCLPERTCEMDKSPVSN----- 711  
 QY 350 RREWNKOLPCSENPRTEDVPIWITLNSIIOKVNEFSSDELLGSDSHGSESNAY 409  
 DB 712 -----GKEKVLVS-----ACLPASVQULSISFESQENLIG--DHNGSTLKLTP 753  
 QY 410 ADVLYDLNEVDYSGSSEKIDLASDPHEALICK--SERVHKSVESNIEDKTFGKT--- 464  
 DB 754 SKRLPL-----SKADWVSREK---MCKMEPEKLOCESKCNIE--LSKNILEY 795  
 QY 465 -----YRKASLPNLSHVTENLIGAFVTEPOIIOGRPLTNKAKRRRPTSGHLPDEFK 519  
 DB 796 NETCITISESKTIPGLPPEBNTIEVASSMKSPNQNAKVIQDKQSP-----FI- 846  
 QY 520 KADLAOKTPEMI--NOGTNOTEONGQVNNITNSGHE-----KTKGDSION----- 564  
 DB 847 -SEVAANMNMSEELFPSPGNNFAQVYNKCKKPDGSSVELQDEEDLSHTQGPISLAKNPMAY 905  
 QY 565 -----EKNPPIESLEK---ESAFKTKA----- 584  
 DB 906 DEDVDNAHAOYLITKDSDSLAVVHDYTEKSRNNIIOKGTEDKPKFSSSLNMMKSDGN 965  
 QY 585 -----EPIS-----STSNMELINTNSKAPK--NRLRKSSTR----- 618  
 DB 966 SDCSDKWSFELDEVLNHNFGSFRTASNKEIKLEHNKSKKMFEDIEOYPTRLACID 1025  
 QY 619 HIAHLELVYSRNLSPNCTELQIDSCSSSEIKKKRYNOMPYRHSNLOLMEGKEPATGA 678  
 DB 1026 IYVTLPLANQKKLSEHIFDLK-----SVTTVSTOSHNOSSVSH----- 1064  
 QY 679 KSNKNNEBOTSKRHSDTPELKLINAPGSGFTKCSNTSELKEFVNDLPREEKEKLETV 738  
 DB 1065 -----EDDTAPOM--LSSKODFHNNLTTSOAKETTELSTLIEESGSOFEPT 1110  
 QY 739 K-----VSNAADEKDLMLSGERYLOT--ERSVSSSSISLVPGDTGYOESISLLEVST 790  
 DB 1111 QRRKPSHIQONTSEVGNOMV-----VLSTASKEMKOTDHLPLVDPSPVSGQDHSKOFEGSA 1166  
 QY 791 LCKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFRY--PLGHEVNHSHETSIEME 848  
 DB 1167 -GVKQSEPHLEBDTCN--KNT-----SCLPLNINEMEFGRICSLGTKLVSNE----- 1212  
 QY 849 ESLELDQYIQTNTFKYSKROSFALFSPNABEECA-----TFSAHSGS-----LKQO 895  
 DB 1213 -----ALRKAMKLFSDIENSEEBSAVGPRGFSSSAHSDSVASVEFKIKRP 1257  
 QY 896 SPKVTFE-----CEEKEENO-----GKN-ESNIRKPVQ-----T 922  
 DB 1258 NTEKSFDEKSSKCOVTLQNNIEMKTTCTIYVGRNPEKTIKNTKHEDSYTSQORNNLENSDGS 1317

QY 923 VNITAGEPVV---GQDKRPVD-NAKCSIKGSRFCLSSQFRGNETGLITENKHLQONPY 978  
 DB 1318 MSSTSG-PVYIHKGDSDLPADQSGKC-----PESC--TQVARENNQIKENISDLCTLEI 1369  
 QY 979 RIRPLFPKSFVYTKCKKNLLENEFEHNSPBEREMENEPSTVSTISNNINRENVFK 1038  
 DB 1370 MKAETCKMKSSDKKOLPSDKMERIKKEFNIS-----FOYASQKNR--VSKE 1414  
 QY 1039 ASSNINVEGSSPNEVGSSINELGSDPENIOAELGRNRGPKLNAMLRGLVLOPEVYKQSL 1098  
 DB 1415 SLKSVNIIFRRETDL-----TVISLNSKILHGIKDKMHTSCHKKAISIKKVEDHF 1469  
 QY 1099 PGSNCKH-PEIKROEYEEVQVNTDFSPYLIS-----DNLEQPMGSSHASQ 1144  
 DB 1470 PIVTSQLPQAQHPRE-----IESRKEPILLFPHNASGKKVIMQESLQK-----YKN 1518  
 QY 1145 VCSETPDOLLDOELIEDTSFAEND--IKSSAVFSKSVQKGLSNPSPTFTHTILAQGY 1202  
 DB 1519 LCGETO-----YVRKTASFQSGSKPLDKSKKELTAYEKIEVT----- 1556  
 QY 1203 RRGAKKLESSEENLSSDEELPCOHLLEFKVNNIPSQSTRHSTVATE----- 1250  
 DB 1557 ---ASKCEQNVSKETEMLPQONTNMTQENTLTSNGTSKVDENIENNEKPRIV 1613  
 QY 1251 CLSKN---TEENILS--LKNSLNCSNOYILAK----- 1278  
 DB 1614 CIGQSSPYTEDSALAVYTEDSKTCVRESSLKGKRWLEQGDKLGTRNTIKIECVKEH 1673  
 QY 1279 -----ASQEH-----LSEETKCSASLSSQCSLEDL----- 1306  
 DB 1674 TEDFAGNASYEHRLVIRTEIDTNHSENOVSTLLSDPNVCHSYLSQSSCHCDMDHNS 1733  
 QY 1307 -----TAN-----TNTQ----- 1313  
 DB 1734 GFPLAKNIDSDVPPDKMNAEGNTISPRVSAFKERNLHPQNTINEXCYOKLETNTNSPANKD 1793  
 QY 1314 ---DPELIGSSK-----QMRHOSQGVGLSDKELVSDDERGTELENNQEPQSD 1362  
 DB 1794 VALDPILDSRNCKVSGSLVFTTAHQETE-----RTKEIYDNCYKIV-----EDNRQ 1841  
 QY 1363 SNLGEAASGCESETSVSEDCSGLSQSDIITTOORDT-----MOHNLIKLOQEMAE 1414  
 DB 1842 SKPDTCOTSCHKVLDSKPFICPSSSGDVCINARKDSFCILNHOIOLHN-----QSMS- 1894  
 QY 1415 EAVLEHGQSPNSYPIIISDSSALDLRNPQ-----STSEKAVLTSQKSSSE 1462  
 DB 1895 -----GLKKAATVPVGLTWTQTSKSTIREPPQAAHPSTRYGIFSTASGKAIVDSASLE 1947  
 QY 1463 YPISQNPBGISADKFEVSADSSSTSKNKEP--GVERSSPKCPSLDRLRWYHSCSGSLQ-- 1518  
 DB 1948 -KARQVFSKMDGAKQLSSVNSLEGNKPRHSAKRENS-----VHSTQGYLSLP 1996  
 QY 1519 -----NRNTP-----SOELIKVYV-----EEOOLEESG-PHDLTE 1549  
 DB 1997 KPLPGWNVSSVFSGSTAGKLVTVESALHKYKGMLEEDLIRTEHTLQHSPIPEVSK 2056  
 QY 1550 TSVLPRQDLEGPY--LESGLSLESDDESPSPDRAPESARVGNINPSTSLKVPOLKA 1608  
 DB 2057 ILPOCAELRTPEYPAVNSKIQKTYNDKSSLPNWK--ESQSSGNTOSIEVSLQSOERN 2114  
 QY 1609 ESAQSPAAHNTDTAGYNMEEVSREKPELTASTERVNRKMS-----VWGLTPPEE 1662  
 DB 2115 QDQQL-----VLGTVSHSKANLLQKEDQTLPRKNIKVTKDEMKTFEDVYKTN 2161  
 QY 1663 MLVYKFAKRNHITLNLITEETHVYMK-----TDAB-----FYCERLKYFLCITAG 1710  
 DB 2162 VGEYYSKESENYFEETEAV--ESAKAMMEDDELDSQTHAKSLEFYCPONETLFSNRTK 2219  
 QY 1711 KMWVSFVWYQSIKEKKMLNEND--FEYRGDVYVNGNHOQPKRARS---ODKRTF---R 1762  
 DB 2220 RGVTVDAVGQPIKSLINEDRILESKGSILT-----PSKSTPDGTVKDSRLFTHHM 2273  
 QY 1763 GLEICCYGPF 1772

Fri Jun 27 10:47:55 2003

us-09-734-672-4.rpr

Page 18

Db 2274 SLEPYTCGRF 2283

Search completed: June 27, 2003, 10:24:30  
Job Time : 83 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 27, 2003, 10:03:36 ; Search time 31 Seconds  
(without alignments)

2492.594 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEQNVINAMOK.....LYQCELDTRYLPQIPHSY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9642	99.9	1863	BRCL_HUMAN	P38398 homo sapien
2	6923	71.7	1878	BRCL_CANFA	Q95153 canis fam1
3	4817	49.9	1812	BRCL_MOUSE	P48754 mus musculu
4	357.5	3.7	3329	BRCL_MOUSE	P97929 mus musculu
5	340	3.5	2869	RBPI_PLAVB	Q00798 plasmodium
6	339	3.5	1957	YD86_SCHPO	Q10411 schizosacch
7	327	3.4	1875	MLP1_YEAST	Q02455 saccharomyc
8	327	3.4	1875	BRCL_HUMAN	P51587 homo sapien
9	325	3.4	3911	AKA9_HUMAN	Q99966 h-a-kinase
10	323.5	3.4	3210	CENF_HUMAN	P49454 homo sapien
11	322	3.3	1658	YMG7_YEAST	Q03661 saccharomyc
12	321.5	3.3	4377	ANK3_HUMAN	Q12955 homo sapien
13	319.5	3.3	2492	ATRX_HUMAN	P46100 homo sapien
14	317.5	3.3	1972	P531_HUMAN	Q12888 homo sapien
15	316	3.3	1727	ALM1_SCHPO	Q9ute5 schizosacch
16	313.5	3.2	1679	Y109_YEAST	P40457 saccharomyc
17	312	3.2	3924	ANK2_HUMAN	Q01484 homo sapien
18	310.5	3.2	2845	APC_MOUSE	Q61315 mus musculu
19	308.5	3.2	2748	NUM1_YEAST	Q00402 saccharomyc
20	308	3.2	2230	GOG4_HUMAN	Q13439 homo sapien
21	307.5	3.2	1790	USO1_YEAST	P25386 saccharomyc
22	304.5	3.2	1539	TANA_XENLA	Q01507 xenopus lae
23	303.5	3.1	1539	Y373_HUMAN	Q13023 homo sapien
24	301.5	3.1	2319	AKA6_HUMAN	P14873 mus musculu
25	301	3.1	2464	MAPB_MOUSE	P25054 homo sapien
26	299.5	3.1	2843	APC_HUMAN	P11055 homo sapien
27	296.5	3.1	1940	MYH3_HUMAN	P12270 homo sapien
28	294.5	3.1	2349	TPR_HUMAN	P46821 homo sapien
29	293	3.0	2468	MAPB_HUMAN	Q02224 homo sapien
30	290	3.0	2663	CENF_HUMAN	Q90953 gallus gall
31	290	3.0	3562	PGCY_CHICK	P08964 saccharomyc
32	288.5	3.0	1928	MTSL_YEAST	Q61687 mus musculu
33	288	3.0	2476	ATRX_MOUSE	

34	286.5	3.0	1781	1	AKAC_HUMAN	Q02952 homo sapien
35	285.5	3.0	2842	1	APC_RAT	P70478 rattus norv
36	284	2.9	3256	1	KI67_HUMAN	P46013 homo sapien
37	282.5	2.9	1940	1	MYH3_RAT	P12847 rattus norv
38	282	2.9	1233	1	YF16_YEAST	P43597 saccharomyc
39	281.5	2.9	1934	1	MYH7_MESAU	P13540 mesocricetu
40	280.5	2.9	1935	1	MYH7_HUMAN	P12883 homo sapien
41	280	2.9	1253	1	DSPB_HUMAN	Q9nzv4 homo sapien
42	278.5	2.9	3358	1	PGCY_MOUSE	Q62059 mus musculu
43	276.5	2.9	1935	1	MYH7_PIG	P79293 sus scrofa
44	276.5	2.9	2459	1	MAPB_RAT	P15205 rattus norv
45	275.5	2.9	1433	1	REST_CHICK	Q42184 gallus gall

## ALIGNMENTS

RESULT 1  
BRCL\_HUMAN STANDARD; PRT; 1863 AA.  
AC P38398;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Breast cancer type 1 susceptibility protein.  
GN BRCA1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT ARG-1775.  
RX MEDLINE=95025896; Pubmed=7545954;  
RA Miki Y., Swensen J., Shattuck-Eidens D., Futreal P.A., Harshman K.,  
RA Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R.,  
RA Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hatlier T.,  
RA Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z.,  
RA Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P.,  
RA Ward J., Tonlin P., Nared S., Bristow P.K., Norris F.H., Helvering L.,  
RA Morrison P., Rostleek P., Lai M., Barrett J.C., Lewis C., Neuhausen S.,  
RA Cannon-Albright L., Goddard D., Wiseman R., Kamb A., Skolnick M.H.;  
RT "A strong candidate for the breast and ovarian cancer susceptibility  
RT gene BRCA1".  
RL Science 266:66-71(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97092865; Pubmed=8938427;  
RA Smith T.M., Lee M.K., Szabo C.I., Jerome N., McEuen M., Taylor M.,  
RA Hood L., King M.-C.;  
RT "Complete genomic sequence and analysis of 117 kb of human DNA  
RT containing the gene BRCA1".  
RL Genome Res. 6:1029-1049(1996).  
RN [3]  
RP IDENTIFICATION OF BRCA1 AS MEMBER OF BASC.  
RX MEDLINE=20245492; Pubmed=10781365;  
RA Wang Y., Cortez D., Yazdi P., Neff N., Ellledge S.J., Qin J.;  
RT "BASC, a super complex of BRCA1-associated proteins involved in the  
RT recognition and repair of aberrant DNA structures".  
RL Genes Dev. 14:927-939(2000).  
RN [4]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=96400954; Pubmed=8807330;  
RA Couch F.J., Weber B.L.;  
RT "Mutations and polymorphisms in the familial early-onset breast  
RT cancer (BRCA1) gene. Breast Cancer Information Core".  
RL Hum. Mutat. 8:8-18(1996).  
RN [5]  
RP VARIANTS LEU-1637; G1U-1708 AND ARG-1775.  
RX MEDLINE=95025878; Pubmed=7939630;  
RA Futreal P.A., Liu Q., Shattuck-Eidens D., Cochran C., Harshman K.,  
RA Tavtigian S., Bennett L.M., Haugen-Strano A., Swensen J., Miki Y.,  
RA Eddington K., McClure M., Frye C., Weaver-Felhaus J., Ding W.,  
RA Gholami Z., Soederkvist P., Terry L., Jhanwar S., Berchuk A.,

RA Iglehart J.D., Marks J., Ballinger D.G., Barrett J.C., Skolnick M.H.,  
 RA Kamb A., Wiseman R.;  
 RT "BRCA1 mutations in primary breast and ovarian carcinomas";  
 RL Science 266:120-122(1994).  
 RN [6]  
 RP VARIANTS BC GLY-64, AND VARIANTS ALA-772; ASN-1040 AND GLY-1443.  
 RX MEDLINE=95201806; PubMed=78944491;  
 RA Castilla L.H., Couch F.J., Erdos M.R., Hoskins K.F., Calzone K.,  
 RA Gaber J.E., Boyd J., Lubin M.B., Deshano M.L., Brody L.C.,  
 RA Collins F.S., Weber B.L.;  
 RT "Mutations in the BRCA1 gene in families with early-onset breast and  
 RT ovarian cancer.";  
 RL Nat. Genet. 8:387-391(1994).  
 RN [17]  
 RP VARIANT BC G-61, AND VARIANTS R-356; G-1038; N-1040; R-1183 AND  
 RP G-1613.  
 RX MEDLINE=95201808; PubMed=78944493;  
 RA Friedman L.S., Ostermeyer E.A., Szabo C.I., Dowd P., Lynch E.D.,  
 RA Rowell S.E., King M.-C.;  
 RT "Confirmation of BRCA1 by analysis of germline mutations linked to  
 RT breast and ovarian cancer in ten families.";  
 RL Nat. Genet. 8:399-404(1994).  
 RN [8]  
 RP VARIANT BC GLY-61.  
 RX MEDLINE=96108965; PubMed=8554067;  
 RA Serova O., Montagna J., Torchard D., Narod S.A., Tonin P., Sylia B.,  
 RA Lynch H.T., Feunteun J., Lenoir G.M.;  
 RT "A high incidence of BRCA1 mutations in 20 breast-ovarian cancer  
 RT families";  
 RL Am. J. Hum. Genet. 58:42-51(1996).  
 RN [9]  
 RP VARIANT BOC TRP-841.  
 RX MEDLINE=97123469; PubMed=8968716;  
 RA Baker D.F., Almeida E.F.A., Casey G., Fain P.R., Liao S.-Y.,  
 RA Masunaka I., Noble B., Kurosaki T., Anton-Culver H.;  
 RT "BRCA1 R841W: a strong candidate for a common mutation with moderate  
 RT phenotype.";  
 RL Genet. Epidemiol. 13:595-604(1996).  
 RN [10]  
 RP VARIANTS BC AND BOC.  
 RX MEDLINE=96372821; PubMed=8776600;  
 RA Durocher F., Shattuck-Eidens D., McClure M., Labrie F.,  
 RA Skolnick M.H., Goltgar D.E., Simard J.;  
 RT "Comparison of BRCA1 polymorphisms, rare sequence variants and/or  
 RT missense mutations in unaffected and breast/ovarian cancer  
 RT populations";  
 RL Hum. Mol. Genet. 5:835-842(1996).  
 RN [11]  
 RP VARIANTS BC MET-271 AND SER-1150.  
 RX MEDLINE=96303704; PubMed=8723683;  
 RA Katagiri T., Eml M., Ito I., Kobayashi K., Yoshimoto M., Iwase T.,  
 RA Kasumi F., Miki Y., Skolnick M.H., Nakamura Y.;  
 RT "Mutations in the BRCA1 gene in Japanese breast cancer patients.";  
 RL Hum. Mutat. 7:334-339(1996).  
 RN [12]  
 RP VARIANT BC GLY-61, AND VARIANTS ARG-239; TRP-841 AND ILE-1512.  
 RX MEDLINE=98430998; PubMed=9760198;  
 RA Dong J., Chang-Claude J., Wu Y., Schumacher V., Debatin I., Tonin P.,  
 RA Royer-Pokora B.;  
 RT "A high proportion of mutations in the BRCA1 gene in German  
 RT breast/ovarian cancer families with clustering of mutations in the 3'  
 RT third of the gene";  
 RL Hum. Genet. 103:154-161(1998).  
 RN [13]  
 RP VARIANT BC GLY-64, AND VARIANTS.  
 RX MEDLINE=98141685; PubMed=9482581;  
 RA Andersen T.I., Eiken H.G., Couch F., Kaada G., Skrede M., Johnsen H.,  
 RA Aloysius T.A., Tveit K.M., Tjanebaerg L., Doerum A., Moeller P.,  
 RA Weber B.L., Boerresen-Dale A.-L.;  
 RT "Constant denaturant gel electrophoresis (CDGE) in BRCA1 mutation  
 RT screening";  
 RL Hum. Mutat. 11:166-174(1998).  
 RN [14]

RP VARIANTS BC S-22; L-461; D-465; V-552; S-892; D-960; I-1025 AND  
 RP A-1047.  
 RX MEDLINE=98272917; PubMed=9609997;  
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaiishi K., Abe R.,  
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Nanda K.,  
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furutsuma Y.,  
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,  
 RA Houga S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,  
 RA Sonoo H., Kurebayashi J.-I., Shimotsu K., Nakamura Y., Miki Y.;  
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in  
 RT Japanese breast cancer families";  
 RL J. Hum. Genet. 43:42-48(1998).  
 RN [15]  
 RP VARIANT OC ARG-1749.  
 RX MEDLINE=20455732; PubMed=10486320;  
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,  
 RA Ponder B.A.J.;  
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial  
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility  
 RT genes";  
 RL Am. J. Hum. Genet. 65:1021-1029(1999).  
 RN [16]  
 RP VARIANT BC S-346, AND VARIANTS L-871; G-1038; R-1183 AND G-1613.  
 RX MEDLINE=99254821; PubMed=10323242;  
 RA Li S.S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,  
 RA Huang H.-W., Chen L.-M., Xao H.-W., Chen J.-H., Tseng J.-N., Chen A.,  
 RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;  
 RT "Molecular characterization of germline mutations in the BRCA1 and  
 RT BRCA2 genes from breast cancer families in Taiwan";  
 RL Hum. Genet. 104:201-204(1999).  
 RN [17]  
 RP VARIANTS OC, AND VARIANTS.  
 RX MEDLINE=99214030; PubMed=10196379;  
 RA Janzic S.A., Zlogas A., Krumroy L.M., Krasner M., Plummer S.J.,  
 RA Cohen P., Gilead M., Barker D., Haile R., Casey G., Anton-Culver H.;  
 RT "Germline BRCA1 alterations in a population-based series of ovarian  
 RT cancer cases";  
 RL Hum. Mol. Genet. 8:889-897(1999).  
 RN [18]  
 RP FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN  
 CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE.  
 CC SUBUNIT: PART OF THE BRCA1-ASSOCIATED GENOME SURVEILLANCE COMPLEX  
 CC (BASC), WHICH CONTAINS BRCA1, MSH2, MSH6, MLH1, ATM, BLM, PMS2 AND  
 CC THE RAD50-MRE11-NBS1 PROTEIN COMPLEX. THIS ASSOCIATION COULD BE A  
 CC DYNAMIC PROCESS CHANGING THROUGHOUT THE CELL CYCLE AND WITHIN  
 CC SUBNUCLEAR DOMAINS. CIP1 INTERACTS SPECIFICALLY WITH THE BRC1  
 CC DOMAINS.  
 CC SUBCELLULAR LOCATION: Nuclear.  
 CC POLYMORPHISM: THERE IS EVIDENCE THAT THE PRESENCE OF THE RARE FORM  
 CC OF Q356R AND L871P POLYMORPHISMS MAY BE ASSOCIATED WITH AN  
 CC INCREASED RISK FOR DEVELOPING OVARIAN CANCER.  
 CC DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,  
 CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE  
 CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF  
 CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-  
 CC ONSET BREAST CANCER. MUTATIONS IN BRCA1 ARE THOUGHT TO BE  
 CC RESPONSIBLE FOR 45% OF INHERITED BREAST CANCER AND MORE THAN 80%  
 CC OF INHERITED BREAST AND OVARIAN CANCER (BOC). MOREOVER, BRCA1  
 CC CARRIERS HAVE A 4-FOLD INCREASED RISK OF COLON CANCER, WHEREAS  
 CC MALE CARRIERS FACE A 3-FOLD INCREASED RISK OF PROSTATE CANCER.  
 CC SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC SIMILARITY: CONTAINS 2 BRC1 DOMAINS.  
 CC DATABASE: NAME=holmolebase; NOTE=BRCA1 entry;  
 CC WWW="http://biominformatics.weizmann.ac.il/holmolebase/entries/brca1.htm".  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U14680; AAA73985.1; -

Query Match 99.9%; Score 9642; DB 1; Length 1863;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MDLSALFVEEYOVYINAMOKILECPICLLEKEPVSTKCDHIFCKFCMLKLNOKKPSQ 60
DB 1 MDLSALFVEEYOVYINAMOKILECPICLLEKEPVSTKCDHIFCKFCMLKLNOKKPSQ 60
QY 61 CPLCKNDITKRSLOESRFQSOIVEELIKITCAPOLODGLYANSYNNAKKNENSPBLKD 120
DB 61 CPLCKNDITKRSLOESRFQSOIVEELIKITCAPOLODGLYANSYNNAKKNENSPBLKD 120
QY 121 EVSLIISGMGYRNAKRLJOSEPEPNSLOETSLVSQSLNGLTVPLRTKORLOPOKTSYVI 180
DB 121 EVSLIISGMGYRNAKRLJOSEPEPNSLOETSLVSQSLNGLTVPLRTKORLOPOKTSYVI 180
QY 181 ELGSDSSEDTYVNAKYCVSGDQELLQITPOGTRDEISLDSAKKAACEFSETDYNTENHQ 240
DB 181 ELGSDSSEDTYVNAKYCVSGDQELLQITPOGTRDEISLDSAKKAACEFSETDYNTENHQ 240
QY 241 PSNNDLNTTEKRAERPEREKYQSSVSNLHVEPCGTTHASSLOHENSLLITKDRANVE 300
DB 241 PSNNDLNTTEKRAERPEREKYQSSVSNLHVEPCGTTHASSLOHENSLLITKDRANVE 300
QY 301 KAERCNKSQOPGLARSOHNMAKSKETCNDRTPTTEKRYDLNADPLCEKKNKOKLPC 360
DB 301 KAERCNKSQOPGLARSOHNMAKSKETCNDRTPTTEKRYDLNADPLCEKKNKOKLPC 360
QY 361 SENPRDTEVPWITLNSIIOKVNEWFSRSDDELLGSDSDSHDESESNKAVADVLVNEVD 420
DB 361 SENPRDTEVPWITLNSIIOKVNEWFSRSDDELLGSDSDSHDESESNKAVADVLVNEVD 420
QY 421 EYSGSSKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKAALPLNLSHTEN 480
DB 421 EYSGSSKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKAALPLNLSHTEN 480
QY 481 LIIGAFVTEPOIIIOERPLTNKLRKRRTSGLHPEDFIKKRADLAVOTPEMINOQTOTE 540
DB 481 LIIGAFVTEPOIIIOERPLTNKLRKRRTSGLHPEDFIKKRADLAVOTPEMINOQTOTE 540
QY 541 QNQNVMNITSSGHENKTGDSIONEKNPNIESIEKSAFTKTAEDPISSISIMNELENT 600
DB 541 QNQNVMNITSSGHENKTGDSIONEKNPNIESIEKSAFTKTAEDPISSISIMNELENT 600
QY 601 HNSKAPKRNLRKRSSTRHIALELVYSRNLSPNCELOIDGSSSEETKRRKKNOMPV 660
DB 601 HNSKAPKRNLRKRSSTRHIALELVYSRNLSPNCELOIDGSSSEETKRRKKNOMPV 660
QY 661 RHSRNLQLMGKBPATGAKKSNKPNEDQTSKRHDSJTFPELKITNAPGSFTKCSNTSELKE 720
DB 661 RHSRNLQLMGKBPATGAKKSNKPNEDQTSKRHDSJTFPELKITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPLEPREKEKLETVKVSNNNAEDPKDMLSGERYLOTERSVESISSILVPGTDYGTQ 780
DB 721 FVNPSPLEPREKEKLETVKVSNNNAEDPKDMLSGERYLOTERSVESISSILVPGTDYGTQ 780
QY 781 ESISLLEVTSLGAKTEPNKCVSOGAFAFENPKGLIHGCSKDNNDPTEGFFYPLGHEVNHNS 840
DB 781 ESISLLEVTSLGAKTEPNKCVSOGAFAFENPKGLIHGCSKDNNDPTEGFFYPLGHEVNHNS 840
QY 841 RETSIEMESELDAOYLONTFKVSKRSQFALFSPNGAEECATFSAHSGSLKKQSPKYT 900
DB 841 RETSIEMESELDAOYLONTFKVSKRSQFALFSPNGAEECATFSAHSGSLKKQSPKYT 900
QY 901 FEBCQKFEENGKKNESNIPQVOTVNTAGFPVYQOKDKPVUNAKCSIKGSRFLSSQFRG 960
DB 901 FEBCQKFEENGKKNESNIPQVOTVNTAGFPVYQOKDKPVUNAKCSIKGSRFLSSQFRG 960
QY 961 NETGLITPNKHGLLONPYRIPLPLPIKSFVTKCKKNLLEENPEHSMSPERMGNNIP 1020
DB 961 NETGLITPNKHGLLONPYRIPLPLPIKSFVTKCKKNLLEENPEHSMSPERMGNNIP 1020
```

```
QY 1021 STVSTISRNINRENVFEKASSNINEVGSSSTNEVGSSINELGSSDENIQAELGNNRPKL 1080
DB 1021 STVSTISRNINRENVFEKASSNINEVGSSSTNEVGSSINELGSSDENIQAELGNNRPKL 1080
QY 1081 NAMIRLCVLOPEVYKOSLPGSNCKNHPETIKKOEVEEVQVYNTDPSPLYISDNLEQPMGSS 1140
DB 1081 NAMIRLCVLOPEVYKOSLPGSNCKNHPETIKKOEVEEVQVYNTDPSPLYISDNLEQPMGSS 1140
QY 1141 HASOVSETPDDLLDDEIEKEDTSEFANDIKESSAVSKSVQKGLSRSPPTHTHLAQ 1200
DB 1141 HASOVSETPDDLLDDEIEKEDTSEFANDIKESSAVSKSVQKGLSRSPPTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSEDEBELPCFOHLLFGKNNIPSOSTRSHYATATCLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSEDEBELPCFOHLLFGKNNIPSOSTRSHYATATCLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVITLAKASQEHHLSEETKCSASLFFSSQSELEDTAANTNOOPFLIGS 1320
DB 1261 LSLKNSLNDCSNOVITLAKASQEHHLSEETKCSASLFFSSQSELEDTAANTNOOPFLIGS 1320
QY 1321 SKQMRHQSSEQVGLSDKELVSDDEERGCTGLEENNOEQQMDSNLGAAGCESFTSVSE 1380
DB 1321 SKQMRHQSSEQVGLSDKELVSDDEERGCTGLEENNOEQQMDSNLGAAGCESFTSVSE 1380
QY 1381 DCSGLSSQSDILATTOOQDRTMOHNLIKLOQMAELEAVLEOHGSOPSNYSYIISDSSALE 1440
DB 1381 DCSGLSSQSDILATTOOQDRTMOHNLIKLOQMAELEAVLEOHGSOPSNYSYIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSNKKEPGEVRSRSPSK 1500
DB 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSNKKEPGEVRSRSPSK 1500
QY 1501 CPSLDDRWYHMSGSLQNNRNPQOEELIKVYVVEEQULESGPHDLTETSYLPRDLEG 1560
DB 1501 CPSLDDRWYHMSGSLQNNRNPQOEELIKVYVVEEQULESGPHDLTETSYLPRDLEG 1560
QY 1561 TPYLESGISLFPDDESPEDRAPESARVGINIPSSSTALKVPOLKVAESAQSPAANHTT 1620
DB 1561 TPYLESGISLFPDDESPEDRAPESARVGINIPSSSTALKVPOLKVAESAQSPAANHTT 1620
QY 1621 DTAGYNAAMESVSRKEKELTASTERVYKRMASVYVSGITPEEFMLYKFAKKHHITLTNLI 1680
DB 1621 DTAGYNAAMESVSRKEKELTASTERVYKRMASVYVSGITPEEFMLYKFAKKHHITLTNLI 1680
QY 1681 TEETTHVYVMTDAEPVCEKTLKFLTAGGKWWVSYFWVQSIKERMLNEHDFEVAGDV 1740
DB 1681 TEETTHVYVMTDAEPVCEKTLKFLTAGGKWWVSYFWVQSIKERMLNEHDFEVAGDV 1740
QY 1741 VNGRNHQPARRAESODRKIFRGLEICCYGPTNMPDQLEMMVYQLCGASVYKELSSFTL 1800
DB 1741 VNGRNHQPARRAESODRKIFRGLEICCYGPTNMPDQLEMMVYQLCGASVYKELSSFTL 1800
QY 1801 GTGVHPILVVQPDAMTEEDNGFHAIGONCEAPVYTRRENVLDLSVALYOCQELDTYLIPOIPH 1860
DB 1801 GTGVHPILVVQPDAMTEEDNGFHAIGONCEAPVYTRRENVLDLSVALYOCQELDTYLIPOIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863
```

RESULT 2  
BRC1\_CANFA  
ID BRC1\_CANFA STANDARD; PRT; 1878 AA.  
AC 095153;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Breast cancer type 1 susceptibility protein homolog.  
GN BRC1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

```

OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97026289; PubMed=8872468;
RA Szabo C.I., Wagner L.A., Francisco L.V., Roach J.C., Argonza R.,
RT "Human, canine and murine BRCA1 genes: sequence comparison among
RT species.";
RL Hum. Mol. Genet. 5:1289-1298(1996).
CC -! FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN
CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
CC SIMILARITY).
CC -! SUBUNIT: CTP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY
CC SIMILARITY).
CC -! SUBCELLULAR LOCATION: Nuclear (Potential).
CC -! SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -! SIMILARITY: CONTAINS 2 BRCT DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U50709; AAC48663.1; -
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR002378; BRCT.
DR InterPro: IPR001841; Znf_Lring.
DR Pfam: PF00097; Zf-C3HC4; 1.
DR Pfam: PF00533; BRCT; 2.
DR PRINTS: PR00493; BRCTCANCER1.
DR SMART: SM00292; BRCT; 2.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00172; BRCT; 2.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR PROSITE: PS50089; ZF_RING_2; 1.
DR Zinc-finger: DNA-binding; Nuclear protein; Anti-oncogene; Repeat.
FT ZN_FING 24 65 RING-TYPE.
FT DOMAIN 1652 1739 BRCT 1.
FT DOMAIN 1764 1863 BRCT 2.
FT SEQUENCE 1878 AA; 208446 MW; 4799865030327C4F CRC64;
SQ

```

Query Match 71.7%; Score 6923; DB 1; Length 1878;  
 Best Local Similarity 74.2%; Pred. No. 2.5e-282;  
 Matches 1390; Conservative 165; Mismatches 301; Indels 18; Gaps 13;

```

OY 1 MDLSLRVEEYVONVINAOKILECPICDELKEPVSTKCDHIFCKFCMLKLNOKKGPQ 60
DB 1 MDLSLRVEEYVONVINAOKILECPICDELKEPVSTKCDHIFCKFCMLKLNOKKGPQ 60
OY 61 CPLKNDITKRSLOSTRPSQLVEELKTIICAFOLDTGLEYANSYNPAKKNENSPHLND 120
DB 61 CPLKNDITKRSLOSTRPSQLVEELKTIICAFOLDTGLEYANSYNPAKKNENSPHLND 120
OY 121 EYSIIOSMGYRNARAKRLQSEPEENPSLOETLSVQLSNLGTVRTLKRIOPKTSYVI 180
DB 121 EYSIIOSMGYRNARAKRLQSEPEENPSLOETLSVQLSNLGTVRTLKRIOPKTSYVI 180
OY 121 EYSIIOSMGYRNARAKRLQSEPEENPSLOETLSVQLSNLGTVRTLKRIOPKTSYVI 179
DB 121 EYSIIOSMGYRNARAKRLQSEPEENPSLOETLSVQLSNLGTVRTLKRIOPKTSYVI 179
OY 181 ELGSSSEEDTVNKATYCSVQDDELQITPOGTRDEISLDSAKKACESEETDVNTNHHQ 240
DB 181 ELGSSSEEDTVNKATYCSVQDDELQITPOGTRDEISLDSAKKACESEETDVNTNHHQ 240
OY 180 ELGSSSEEDTVNKATYCSVQDDELQITPOGTRDEISLDSAKKACESEETDVNTNHHQ 237
DB 180 ELGSSSEEDTVNKATYCSVQDDELQITPOGTRDEISLDSAKKACESEETDVNTNHHQ 237
OY 241 PSNNLNTTERKAARHPEKYGSSVSNLHVEPCGNTHAASSLOHENSLLTKKDMNVE 300
DB 241 PSNNLNTTERKAARHPEKYGSSVSNLHVEPCGNTHAASSLOHENSLLTKKDMNVE 300
OY 238 SSNKLUTTEKHAHKHPEKYGISVSNLHVEPCGNTHAASSLOHENSLLTKKDMNVE 297
DB 238 SSNKLUTTEKHAHKHPEKYGISVSNLHVEPCGNTHAASSLOHENSLLTKKDMNVE 297
OY 301 KAEFCNKSQOPGLANSQHRNAGSKETCDNRTPSTKEKKYDLNADPCLCEKKEWNKOLPC 360
DB 301 KAEFCNKSQOPGLANSQHRNAGSKETCDNRTPSTKEKKYDLNADPCLCEKKEWNKOLPC 360
OY 298 KAEICNNSQOPGLANSQHRNAGSKETCDNRTPSTKEKKYDLNADPCLCEKKEWNKOLPC 357
DB 298 KAEICNNSQOPGLANSQHRNAGSKETCDNRTPSTKEKKYDLNADPCLCEKKEWNKOLPC 357
OY 361 SENPDTEDEVPIITLNSSTIQKYNEMFSRDELLGSDSDSHDGESNAKAVADLVLENDV 420
DB 361 SENPDTEDEVPIITLNSSTIQKYNEMFSRDELLGSDSDSHDGESNAKAVADLVLENDV 420

```

```

DB 358 SDSPRDSDVPWITLNSSTIQKYNEMFSRDELLGSDSDSHDGESNAKAVADLVLENDV 417
OY 421 EYSGSEKIDLDLADPHEALICKSERVHSKSVESNEDKIECKTKRKAASLPNLSHVEN 480
DB 418 EYSGSEKIDLDLADPHEALICKSERVHSKSVESNEDKIECKTKRKAASLPNLSHVEN 477
OY 481 LIIGAFVTEPOLIIEPRLTNKLRKRRTSGLHPEDFIKADLA-VOKTPREINOGTNOT 539
DB 478 LIIGAFVTEPOLIIEPRLTNKLRKRRTSGLHPEDFIKADLA-VOKTPREINOGTNOT 537
OY 540 EONGVYNTNSGHEKTKGDSIQNEKNPNLIESLEKESAFKXKAPLPISSISNNELELN 599
DB 538 KRGHVIINTNNGPENETEGDYQKEKNANPESLEKESAFKXKAPLPISSISNNELELN 597
OY 600 IHSKAPKRNLRKRSSTHIALELVYARNLSPNCPTELOIDSSSESEIKKKKNOMP 659
DB 598 SSSSKAPKRNLRKRSSTHIALELVYARNLSPNCPTELOIDSSSESEIKKKKNOMP 657
OY 660 VRHSRNLQMECKEPATGAKKSNKPNQTSKRHSDTPPELKTALNAPGFTKCSNTSELK 719
DB 658 VRHNKTLOLMQCKEPATGAKKSNKPNQTSKRHSDTPPELKTALNAPGFTKCSNTSELK 717
OY 720 EYVNSPLPREKEEKLTYKVSNNABDPKDLMSGRVLTQERSVSSISLVPTDYGT 779
DB 718 ECINPGLRREELIESRMTQVSDPDRPKELVLSGRTQTERSVESSTISLVPTDYGT 777
OY 780 QESISLLEVSTIGKATPEPKKCSOCAFENKGLIHGCSKNRNDTEGFKPLGHEVNH 839
DB 778 QDSISLLEVSTIGKATPEPKKCSOCAFENKGLIHGCSKNRNDTEGFKPLGHEVNH 837
OY 840 SRETSIEMESELDAQYLONTFRKVSROSFALESNGNAEBCATFSAHSGSLKROSPKV 899
DB 838 TQETSIEMESELDAQYLONTFRKVSROSFALESNGNAEBCATFSAHSGSLKROSPKV 897
OY 900 TFECEKERNQKESNENIKPVQTVNTTACFPVYQK-DKPVUNACSTIKGSRFLSSQF 958
DB 898 TFECEKERNQKESNENIKPVQTVNTTACFPVYQK-DKPVUNACSTIKGSRFLSSQF 957
OY 959 KGNETGLIPNKHGLNLOPYRIPLPEPKISFKTKCKKMLLENPEEHSMSPEREKGNE 1018
DB 958 KGNETGLIPNKHGLNLOPYRIPLPEPKISFKTKCKKMLLENPEEHSMSPEREKGNE 1017
OY 1019 -IPSTVSTISRNRIKRNVEKSSNINENEGSSTNEVGSSINEIGSSDENIOALGRNRG 1077
DB 1018 -IPSTVSTISRNRIKRNVEKSSNINENEGSSTNEVGSSINEIGSSDENIOALGRNRG 1077
OY 1078 PKLNAMLRGLQPEYVKOSLPGSNCKHPETIKQ-EYEVVQTVNTDPSYLISSNLEOP 1136
DB 1078 PKLNAMLRGLQPEYVKOSLPGSNCKHPETIKQ-EYEVVQTVNTDPSYLISSNLEOP 1137
OY 1137 MGSSHASOYCSFPPDILLDDGELKEDTSFAENDIKESSAVFSVQKGEISRPPTHT 1196
DB 1138 MGSSHASOYCSFPPDILLDDGELKEDTSFAENDIKESSAVFSVQKGEISRPPTHT 1197
OY 1197 HLAQGYRRGAKKLESSEENLSEDELPFCFHLRFGKVNIPQSOSTRSTVATECLSKNT 1256
DB 1198 HLAQGYRRGAKKLESSEENLSEDELPFCFHLRFGKVNIPQSOSTRSTVATECLSKNT 1257
OY 1257 EENLISLAKNSLDCSNQVILAAASOBHILSEETKCSASLFSQSGCELEDLTANTYTDPF 1316
DB 1258 EENLISLAKNSLDCSNQVILAAASOBHILSEETKCSASLFSQSGCELEDLTANTYTDPF 1317
OY 1317 LI--GSSKQMRHOSQGVGLDKELYS--DDEERGCTGLEENN-OEOQSDMSNLGFAAGG 1372
DB 1318 SMDPTSKOYRHOSELDV-LNDKELVSDDDDEREPGLEBDSPOQESYVSDLGFEVAGY 1376
OY 1373 ESETSVSEDCSGLSQSDITLTQOQRTMOHNLKIQOEMAELEAVLEOHGSPNSYPSI 1432
DB 1372 ESETSVSEDCSGLSQSDITLTQOQRTMOHNLKIQOEMAELEAVLEOHGSPNSYPSI 1436
OY 1433 ISDSSALEDLRNPBOSTSKAVLTSOKSSEYFISQNPGLSADKEVYADSDSTSKNNEPG 1492
DB 1433 ISDSSALEDLRNPBOSTSKAVLTSOKSSEYFISQNPGLSADKEVYADSDSTSKNNEPG 1492

```

Db 1437 IADSCSPEDLLNPEONASER-VITSEKSDSPISQNPESLSTDKFOVFLDSTSKNCEPG 1495  
 QY 1493 VERSSPKCPSLDROWMHCSCSGLQNRNYPQOELIKVYDVERQOLESGPHDLTTSY 1552  
 Db 1496 MINSPPQSRLDTRMWHSCPSKSLQDTPNCPQOELIKVYDVERQOLESGPHDLTTSY 1555  
 QY 1553 LPRODLGFTYLLSGLISLFSDDPESDPEAPASARVGNIPSSSTALKYPOKVAESAQ 1612  
 Db 1556 LSRDLEGARYLESGISLFSDDPESDPSHRASLAVSSMPTSTALKYPOKVAESAQ 1615  
 QY 1613 SPAAHTTDTAGVNAMEESVSRKPELSTERYNKMMSVQSLPDEEMLYKFKARKH 1672  
 Db 1616 STAAVHTASTAGYKSSDSYIEKPEVISTRGVNRKISWASGLPKEEMLYKFKARKH 1675  
 QY 1673 HITLNTLITEETHVVKTAEPFCECTLYFLGSIAGKKNVSVFWTOSIKERKMLNEH 1732  
 Db 1676 HISTLNLISEETHVVKTAEPFCECTLYFLGSIAGKKNVSVFWTOSIKERKMLNEH 1735  
 QY 1733 DFEVGRGVNVRNHOQPKRA-----RESODRKIFRGLEICCYGFTNMPDQLEMYQLC 1787  
 Db 1736 DFEVGRGVNVRNHOQPKRARRESQDRSKIFRGLEICCYGFTNMPDQLEMYQLC 1795  
 QY 1788 GASVYKSLSTFTGTCVHPVYVQPDAMTENGHFGMCAPVYTRRWVLDVALYQC 1847  
 Db 1796 GASVYKSLSTFTGTCVHPVYVQPDAMTENGHFGMCAPVYTRRWVLDVALYQC 1855  
 QY 1848 QELDTYLIPOIPHS 1861  
 Db 1856 QELDTYLIPOIPRT 1869

RESULT 3  
 BRCL\_MOUSE STANDARD: PRT: 1812 AA.  
 ID BRCL\_MOUSE P48734; Q60957; Q60983;  
 AC P48734; Q60957; Q60983;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Breast cancer type 1 susceptibility protein homolog.  
 GN BRCA1.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=96177659; PubMed=8634697;  
 RA Abel K.J., Xu J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;  
 RT "Mouse Brca1: localization sequence analysis and identification of  
 evolutionarily conserved domains";  
 RL Hum. Mol. Genet. 4:2265-2273(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=96177660; PubMed=8634698;  
 RA Shaaran S.K., Wims M., Bradley A.;  
 RT "Murine Brca1: sequence and significance for human missense  
 mutations";  
 RL Hum. Mol. Genet. 4:2275-2278(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ;  
 RX MEDLINE=96121367; PubMed=8575748;  
 RA Bennett L.M., Haugen-Striano A., Cochran C., Brownlee H.A.,  
 \*Fiedorek F.T., Jr., Wiseman R.W.;  
 RT "Isolation of the mouse homologue of BRCA1 and genetic mapping to  
 mouse chromosome 11";  
 RL Genomics 29:576-581(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ; TISSUE=Embryo;  
 RX MEDLINE=96067162; PubMed=7590247;

RA Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;  
 RT "Expression of Brca1 is associated with terminal differentiation of  
 ectodermally and mesodermally derived tissues in mice";  
 RL Genes Dev. 9:2712-2722(1995).  
 RN [5]  
 RP SEQUENCE OF 727-1111 FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=96021028; PubMed=7550308;  
 RA Margulis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,  
 RA Abel K.J., Weber B.L., Chodosh L.A.;  
 RT "The developmental pattern of Brca1 expression implies a role in  
 differentiation of the breast and other tissues";  
 RL Nat. Genet. 11:17-26(1995).  
 RN [6]  
 RP SEQUENCE OF 789-1250 FROM N.A.  
 RC STRAIN=129/SVJ;  
 RX MEDLINE=96163506; PubMed=8566965;  
 RA Schroeder E., Badger P., Larson D., Erdos M., Wynshaw-Boris A.,  
 RA Red T., Brody L.;  
 RT "The murine homologue of the human breast and ovarian cancer  
 susceptibility gene Brca1 maps to mouse chromosome 11D";  
 RL Hum. Genet. 97:256-259(1996).  
 CC -1- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN  
 TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY  
 SIMILARITY).  
 CC -1- SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSION IN OTIC VESICLES AT  
 DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS  
 ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN  
 DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS  
 14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN  
 THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,  
 EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF  
 THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO  
 EXPRESSION IN HEART, LIVER OR LUNG.  
 CC -1- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES  
 DRAMATICALLY DURING PREGNANCY. LEVELS FALL DURING LACTATION AND  
 INCREASE AGAIN DURING POST-LACTATIONAL REGRESSION OF THE  
 MAMMARY GLAND.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC -1- SIMILARITY: CONTAINS 2 BRCT DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U35641; AAB17113.1; -  
 DR EMBL: U31625; AAB17114.1; -  
 DR EMBL: U32446; AAA96393.1; -  
 DR EMBL: U36475; AAC52323.1; -  
 DR EMBL: U33835; AAA99742.1; -  
 DR MGD: MGI:104537; Brca1.  
 DR InterPro: IPR001357; BRCT.  
 DR InterPro: IPR002378; BRCT.  
 DR InterPro: IPR001841; Znf\_fing.  
 DR Pfam: PF00097; zf-C3HC4; 1.  
 DR Pfam: PF00533; BRCT; 2.  
 DR PRINTS: PR00493; BRSTCANCER1.  
 DR SMART: SM00292; BRCT; 2.  
 DR SMART: SM00184; RING; 1.  
 DR PROSITE: PS50172; BRCT; 2.  
 DR PROSITE: PS00518; ZF\_RING\_1; 1.  
 DR PROSITE: PS50089; ZF\_RING\_2; 1.  
 DR Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene; Repeat.  
 KW ZN\_FING 24 65  
 FT DOMAIN 1585 1679 BRCT 1.  
 FT DOMAIN 1698 1797 BRCT 2.

ET VARIANT 93 93 F -> L (IN STRAIN 129/SVJ) .  
ET VARIANT 305 305 T -> S (IN STRAIN 129/SVJ) .  
ET VARIANT 319 319 P -> A (IN STRAIN 129/SVJ) .  
ET VARIANT 377 377 Q -> E (IN STRAIN 129/SVJ) .  
ET VARIANT 550 550 K -> Q (IN STRAIN 129/SVJ) .  
ET VARIANT 652 652 A -> P (IN STRAIN 129/SVJ) .  
ET VARIANT 765 765 S -> P (IN STRAIN 129/SVJ) .  
ET VARIANT 917 917 P -> L (IN STRAIN 129/SVJ) .  
ET VARIANT 933 933 C -> S (IN STRAIN 129/SVJ) .  
ET VARIANT 1122 1122 K -> I (IN STRAIN 129/SVJ) .  
ET VARIANT 1206 1206 S -> R (IN STRAIN 129/SVJ) .  
ET VARIANT 1212 1212 RM -> GI (IN STRAIN 129/SVJ) .  
ET VARIANT 1255 1255 S -> R (IN STRAIN 129/SVJ) .  
ET VARIANT 1261 1261 H -> N (IN STRAIN 129/SVJ) .  
ET VARIANT 1264 1264 V -> A (IN STRAIN 129/SVJ) .  
ET VARIANT 1269 1269 P -> A (IN STRAIN 129/SVJ) .  
ET VARIANT 1283 1283 T -> K (IN STRAIN 129/SVJ) .  
ET VARIANT 1337 1337 T -> N (IN STRAIN 129/SVJ) .  
ET VARIANT 1349 1349 P -> T (IN STRAIN 129/SVJ) .  
ET VARIANT 1352 1352 EG -> OR (IN STRAIN 129/SVJ) .  
ET VARIANT 1353 1353 S -> P (IN STRAIN 129/SVJ) .  
ET VARIANT 1381 1381 G -> A (IN STRAIN 129/SVJ) .  
ET VARIANT 1390 1390 V -> D (IN STRAIN 129/SVJ) .  
ET VARIANT 1400 1400 V -> D (IN STRAIN 129/SVJ) .  
ET VARIANT 1503 1503 E -> Q (IN STRAIN 129/SVJ) .  
ET VARIANT 1549 1549 V -> Q (IN STRAIN 129/SVJ) .  
ET VARIANT 1680 1680 T -> K (IN STRAIN 129/SVJ) .  
ET VARIANT 1712 1712 D -> E (IN STRAIN 129/SVJ) .  
ET VARIANT 1721 1721 D -> E (IN STRAIN 129/SVJ) .  
SQ SEQUENCE 1812 AA: 198669 MM: 2291EA74150B86A CRC64:

Query Match 49.9% Score 4817; DB 1; Length 1812;  
Best Local Similarity 55.5%; Pred. No. 2,9e-194;  
Matches 1035; Conservative 261; Mismatches 498; Indels 70; Gaps 32;

QY 1 MDLSLRVEVONVNNAMOKILECPICELIKEPSTKCDHIFCKRCMKLNLQKKRPSQ 60  
DB 1 MDLSVQIOEVONVNHAMOKILECPICELIKEPSTKCDHIFCKRCMKLNLQKKRPSQ 60  
QY 61 CPLCNDITKRSLOESTRFSQLEVEELKICAFOLDTGLEAVANSYFAKKENNSPHELD 120  
DB 61 CPLCNEITKRSLOESTRFSQLEVEELKICAFOLDTGLEAVANSYFAKKENNSPHELD 120  
QY 121 EVSITIQMGYNNRAKRLLOSEPNDLSLOETSLVSLOVNTLGTVTLTKRKORIPQKTSVYI 180  
DB 121 EASITQGYVYNNRRLPQVEBPNGATLKD-SLGVOLSLNGIYVSYKKNROTQPKRKSXYI 179  
QY 181 ELGSSSEDTYKATYCSVGOELIQTPOGTREDESLDSAKKAACEFSEDTVTNHEHO 240  
DB 180 ELDSOSSSEETVTKPGDCSVRDEQLQTAPOEAGDEBKLHSAEBAACEFSE-GIRNLEHO 238  
QY 241 PSNNDLNTTEKRAAERHPEKYGSSVSNLHVEPCGTNTHASSLOHENSLLLTCKDRMNYE 300  
DB 241 CS-DCLPNTENHATERHPEKCSISISNVCBPCGDHAHASSLOPETSSLLLTEDRMNNE 297  
QY 301 KAEFCNKSKOPGLASQHNRAAGSKETCDKDRTPSTIEKKVLDNLADPLCERKENKOKLPC 360  
DB 301 KAEFCNKSKOPGLASQHNRAAGSKETCDKDRTPSTIEKKVLDNLADPLCERKENKOKLPC 360  
QY 298 KAEFCNKSKOPGLASQHNRAAGSKETCDKDRTPSTIEKKVLDNLADPLCERKENKOKLPC 357  
DB 298 KAEFCNKSKOPGLASQHNRAAGSKETCDKDRTPSTIEKKVLDNLADPLCERKENKOKLPC 357  
QY 361 SENPDTEDEVPIITLNNSTIOKYNEMFSRDELLGSDSHDGESESNAAKVADYVDNVEYD 420  
DB 361 SENPDTEDEVPIITLNNSTIOKYNEMFSRDELLGSDSHDGESESNAAKVADYVDNVEYD 420  
QY 358 PENSATITDVPWITLNNSTIOKYNEMFSRDELLGSDSHDGESESNAAKVADYVDNVEYD 417  
DB 358 PENSATITDVPWITLNNSTIOKYNEMFSRDELLGSDSHDGESESNAAKVADYVDNVEYD 417  
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSYESNIEDKIFGTYRKKASLPNLVHTEN 480  
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSYESNIEDKIFGTYRKKASLPNLVHTEN 480  
QY 418 GGFSSSRKTDVTPPHHTLMCKSGRDPFKVEDNTSDKIFGYSYRKGRKSRPHLNVTE 476  
DB 418 GGFSSSRKTDVTPPHHTLMCKSGRDPFKVEDNTSDKIFGYSYRKGRKSRPHLNVTE 476  
QY 481 LITCAFVIEPOTIIOERPLTNKLKRRRPPSGLHPEDFIKKADIA-VQKTPREMIINOCT 539  
DB 481 LITCAFVIEPOTIIOERPLTNKLKRRRPPSGLHPEDFIKKADIA-VQKTPREMIINOCT 539  
QY 477 -IIGFTIEPOTIIOERPLTNKLKRRRPPSGLHPEDFIKKADIA-VQKTPREMIINOCT 533  
DB 477 -IIGFTIEPOTIIOERPLTNKLKRRRPPSGLHPEDFIKKADIA-VQKTPREMIINOCT 533  
QY 540 EONGVNMITNSGCHENKTKGDSIONENKNNPIESLEKESAFKTKAPPISSISNMLEJLN 599  
DB 540 EONGVNMITNSGCHENKTKGDSIONENKNNPIESLEKESAFKTKAPPISSISNMLEJLN 599  
QY 534 EPNEDAVSTTSKQENKTAGSNLQEKSAHPTESLKEPEASTAGAKSISNSVSDLEVEJLN 593  
DB 534 EPNEDAVSTTSKQENKTAGSNLQEKSAHPTESLKEPEASTAGAKSISNSVSDLEVEJLN 593

QY 600 IHNSKAPKKNRLRKSSTHIALHLELVSRNLSPPNCELTQDSCSSSEIEIKKKKNOMP 659  
DB 594 VHSKAPKKNRLRKSSTHIALHLELVSRNLSPPNCELTQDSCSSSEIEIKKKKNOMP 652  
QY 660 VHSKAPKKNRLRKSSTHIALHLELVSRNLSPPNCELTQDSCSSSEIEIKKKKNOMP 719  
DB 653 VHSKAPKKNRLRKSSTHIALHLELVSRNLSPPNCELTQDSCSSSEIEIKKKKNOMP 711  
QY 720 EPNPSPLEPEEKETLYKVSNNMADPDLMLSGE-RVLQTERVESSSTISLYNCTYX 778  
DB 712 GPNPSPLEPEEKETLYKVSNNMADPDLMLSGE-RVLQTERVESSSTISLYNCTYX 770  
QY 779 TQESILLEVSTLGGKARTPNKCVSQAFAENPKGLIHGCKDNNDTEGKPYPLGHEVN 838  
DB 771 TQESILLEVSTLGGKARTPNKCVSQAFAENPKGLIHGCKDNNDTEGKPYPLGHEVN 829  
QY 839 HSRRETSIEBESELDQOYLTQNTFKYSKRSQFALFSPGNAEBECATFSAHSGSLKQSPK 898  
DB 830 LSOE-KVEMEDSELDQOYLTQNTFKYSKRSQFALFSPGNAEBECATFSAHSGSLKQSPK 884  
QY 899 VPFCEKKEENCKNSNKKPVQYNTITAGPVVGOKKPVDMNCSITKGSRFCLSSQF 958  
DB 885 VPFCEKKEENCKNSNKKPVQYNTITAGPVVGOKKPVDMNCSITKGSRFCLSSQF 941  
QY 959 RGNETGLTPNKHGLLONPYRIPPLPIKSFYKTKCKKNLLENEFEHSMSPEREMNEN 1018  
DB 942 RGNETGLTPNKHGLLONPYRIPPLPIKSFYKTKCKKNLLENEFEHSMSPEREMNEN 1001  
QY 1019 I-PTVSTISRNINRENVKESASSNINEVSGSSTNEIGSSSDENIOALGKRNK 1077  
DB 1002 ILOSTHTVTSLLN-RGNACQEAQS-----GSHIEVCTSDSPFGQGLGRNK 1046  
QY 1078 PKTNAMLRGLVLOPEYKOSLPGSNCKHPEIKKOEVEEYQVNTDFPYLISDLDEPM 1137  
DB 1047 PKTNAMLRGLVLOPEYKOSLPGSNCKHPEIKKOEVEEYQVNTDFPYLISDLDEPM 1101  
QY 1138 GSSHASQVSETPDDLLDDEIKEDTSPFAENDIKSSAFKSVQKGLSPSPFFPTH 1197  
DB 1102 GSSHASQVSETPDDLLDDEIKEDTSPFAENDIKSSAFKSVQKGLSPSPFFPTH 1160  
QY 1198 LAOGRGAKKULESSEENISSEDEBLPCFOHLLPFKVNNTIPQSTHSTVATECLSKTE 1257  
DB 1161 KSOQLTHRASARKLESSEEDSTEDDELPFOHLL-RISTNP-ELTRCSAYATQORPEAE 1218  
QY 1258 ENLSLKNLNCNSOVILAKASOEHHLSSEFKCASLFFSSQCSLEDTANTNTQDFEL 1317  
DB 1219 GTOAPKKGSSSDCNNEVITIEASQEHQPSDEPRGSGMPSOHSAAVOGSTRIANSQDSNF 1278  
QY 1318 IGSSKQMRHQSSEQVGLSDKELVSDDEBGTGLENNQEOQSDNLSGEA-ASGCESET 1376  
DB 1279 IGSSKQMRHQSSEQVGLSDKELVSDDEBGTGLENNQEOQSDNLSGEA-ASGCESET 1336  
QY 1377 SVSESGSLSOSDITLTQOQRTQOHNLIKLOOEAHLEAVLEQKQSGSPNSYPSIIDS 1436  
DB 1337 SVSESGSLSOSDITLTQOQRTQOHNLIKLOOEAHLEAVLEQKQSGSPNSYPSIIDS 1392  
QY 1437 SALEDLRNEQSTSEKAVLTISQSSSEYPIQNPBELSADKEVVS-ADSTSKNKEPVEYR 1495  
DB 1433 SALEDLRNEQSTSEKAVLTISQSSSEYPIQNPBELSADKEVVS-ADSTSKNKEPVEYR 1408  
QY 1496 SSPSKCPSLDDRWYHMSGSLQNNRNPQOELLKVVAVDEEQOLESQPHDLETSTYLP 1555  
DB 1453 SSPSKCPSLDDRWYHMSGSLQNNRNPQOELLKVVAVDEEQOLESQPHDLETSTYLP 1508  
QY 1556 QDEGTPYLESISTLSPD-DEPSDPSEDRAPASAVNGNIPSTSLAKYPLQKVAESAQSP 1614  
DB 1509 QDEGTPYLESISTLSPD-DEPSDPSEDRAPAPASAVNGNIPSTSLAKYPLQKVAESAQSP 1564  
QY 1615 AAHTTDTAGYNAMESVSRKPELTASTERYNKKMSVVGSLPDEEMLYTKARKKHI 1674  
DB 1565 AAHTTDTAGYNAMESVSRKPELTASTERYNKKMSVVGSLPDEEMLYTKARKKHI 1617

QY 1675 TLNLTTEETHVYMTDAEFCVCEKTLKFLGAGKMWVSFYWTQSIKERKMLNEDH 1734  
 Db 1618 TLTDATTEETHYITITDAEFVCEKTLKFLGAGKMWVSFYWTQSIKERKMLNEDH 1677  
 QY 1735 EVGDVAVNGRHHGPPKRRARESDORKEIFRGELICCYGPTFMPPDOLEMMVQLGASVYKE 1794  
 Db 1678 EVGDVAVNGRHHGPPKRRARESDORKEIFRGELICCYGPTFMPPDOLEMMVQLGASVYKE 1736  
 QY 1795 LASFETLGTVHPIVYVOPDAMTEDENGFHAIGMCEAPVYTRRWVLDLSVALYCCOEDTYL 1854  
 Db 1737 LPSLTHDTGALHLYVIVQPSAWTEDESNCPDLCGLKARLVMMQVMDVLDLSISYRCRDAYL 1796  
 QY 1855 IPQI 1858  
 Db 1797 VQNI 1800

RESULT 4  
 BRC2\_MOUSE STANDARD; .PRT; 3329 AA.  
 AC P97929; P97383; 035922;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Breast cancer type 2 susceptibility protein.  
 GN BRCA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=97217789; PubMed=9063750;  
 RA Connor F., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,  
 RA Tait T.M., Freeman T., Ashworth A.;  
 RT Cloning, chromosomal mapping and expression pattern of the mouse  
 RT Brca2 gene.";  
 RL Hum. Mol. Genet. 6:291-300(1997).  
 RN (12)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=97237041; PubMed=9119389;  
 RA Sharon S.K., Bradley A.;  
 RT Murine Brca2: sequence, map position, and expression pattern.";  
 RL Genomics 40:234-241(1997).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=97384941; PubMed=9242436;  
 RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,  
 RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;  
 RT Characterization of the rat and mouse homologues of the BRCA2 breast  
 RT cancer susceptibility gene.";  
 RL Cancer Res. 57:3121-3125(1997).  
 RN (4)  
 RP SEQUENCE OF 18-200 FROM N.A.  
 RX MEDLINE=97075121; PubMed=8917547;  
 RA Rajan J.V., Wang M., Marguis S.T., Chodosh L.A.;  
 RT Brca2 is coordinately regulated with Brca1 during proliferation and  
 RT differentiation in mammary epithelial cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).  
 RN (5)  
 RP SEQUENCE OF 569-625 FROM N.A.  
 RX MEDLINE=97341126; PubMed=9196008;  
 RA McAllister K.A., Ramachandran S., Haugen-Strano A., Fiedorek F.T. Jr.,  
 RA Wiseman R.W.;  
 RT Genetic mapping of the Brca2 breast cancer susceptibility gene on  
 RT mouse chromosome 5.";  
 RL Mamm. Genome 8:540-541(1997).  
 CC -I- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE  
 CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS  
 CC RECOMBINATION (BY SIMILARITY).

CC -I- SUBUNIT: INTERACTS WITH RAD51 (BY SIMILARITY).  
 CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN  
 CC CEREBELLUM, TESTIS, ILEUM, APPENDIX, EPIDIDYMIS, OVARY AND MAMMARY  
 CC GLAND. NO EXPRESSION IN LUNG.  
 CC -I- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES  
 CC DRASTICALLY DURING PREGNANCY.  
 CC -I- SIMILARITY: CONTAINS 7 BRCA2 REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: U82270; AAB48306.1; -;  
 CC EMBL: U72947; AAB40720.1; -;  
 CC EMBL: U65594; AAC23702.1; -;  
 CC EMBL: U89652; AAB71377.1; -;  
 CC EMBL: U89503; AAC53276.1; -;  
 CC MGD: MGI:109337; Brca2.  
 CC InterPro: IPR002093; BRCA2\_repeat.  
 CC Pfam: PF00634; BRCA2; 7.  
 CC PROSITE: PS50138; BRCA2\_REPEAT; 6.  
 CC KW Polymorphism; Repeat.  
 CC REPEAT 981 1015  
 CC REPEAT 1192 1226  
 CC REPEAT 1394 1428  
 CC REPEAT 1491 1525  
 CC REPEAT 1623 1657  
 CC REPEAT 1924 1958  
 CC REPEAT 2004 2038  
 CC REPEAT 44 44  
 CC REPEAT 340 340  
 CC REPEAT 377 377  
 CC REPEAT 407 407  
 CC REPEAT 661 661  
 CC REPEAT 739 739  
 CC REPEAT 1038 1038  
 CC REPEAT 1198 1199  
 CC REPEAT 1257 1257  
 CC REPEAT 1392 1392  
 CC REPEAT 1520 1521  
 CC REPEAT 1583 1583  
 CC REPEAT 1613 1613  
 CC REPEAT 1686 1686  
 CC REPEAT 1799 1799  
 CC REPEAT 1881 1881  
 CC REPEAT 1894 1894  
 CC REPEAT 2141 2141  
 CC REPEAT 2392 2392  
 CC REPEAT 2605 2605  
 CC REPEAT 2648 2648  
 CC REPEAT 2717 2717  
 CC REPEAT 2729 2729  
 CC REPEAT 2814 2814  
 CC REPEAT 2827 2827  
 CC REPEAT 2907 2907  
 CC REPEAT 2929 2929  
 CC REPEAT 3058 3058  
 CC REPEAT 3071 3071  
 CC REPEAT 3081 3081  
 CC REPEAT 3089 3089  
 CC REPEAT 3105 3109  
 CC REPEAT 3220 3220  
 CC REPEAT 3238 3238  
 CC REPEAT 3243 3243  
 CC REPEAT 3245 3245  
 CC SO SEQUENCE 3329 AA; 370658 MM; AAB269APBF05A5D1 CRC64;

Query Match 3.7%; Score 357.5; DB 1; Length 3329;  
 Best Local Similarity 19.2%; Pred. No. 2.2e-07;





AC 000798;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 1 precursor.  
 GN RBP1.  
 OS Plasmodium vivax (strain Belam).  
 OC Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92315338; PubMed=1617731;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 merozoites";  
 RL Cell 69:1213-1226(1992).  
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC HUMAN RETICULOCYTE CELLS.  
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M88097; AAA29743.1;  
 KW Malaria; Receptor; Signal; Transmembrane.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
 FT DOMAIN 18 2807 EXTRACELLULAR.  
 FT TRANSMEM 2808 2826 POTENTIAL.  
 FT DOMAIN 2827 2869 CYTOPLASMIC.  
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;  
 Query Match 3.58; Score 340; DB 1; Length 2869;  
 Best Local Similarity 17.78; Pred. No. 9.9e-07;  
 Matches 384; Conservative 340; Mismatches 742; Indels 704; Gaps 92;  
 QY 1 MDLSARVEE-VONVINAOKILEPCICE-----LIKEPSTCKDHT 42  
 DB 626 IDSLSOTVEKFEYKELIDSKREKIELKEPEKSVTEINRLDGDGASADLHEQIKELIDK 685  
 QY 43 ECKFCMLKLLNOKRGSPGCPPLCKNDITKRSLOSTRFSSQVLVELLKITICAFOLDGLEYA 102  
 DB 686 AKKVHLKELLSLKS-----SVYFTEW-NELNMT-----ASYD 719  
 QY 103 NSNVFAKKENSPENHLKDEVSIISGMGYNNRAKRLLOSEPENSLOETSLISVOLSMGLY 162  
 DB 720 NMEGFSAKK---EKADNDINALYNSVYR-----EDINALIEVEKEVTEENKESLT 766  
 QY 163 RLTLRTKRIOTPOKTSYVIELGSDSEDTYKATYCSVGOELLQI-----TQCTR 213  
 DB 767 EMKADKE-----MEKILQDAKETAKLNF--VSDKLLDYTKKSAEYTNAGLTK 814  
 QY 214 DEIS---LDSAKKACEFSETDYTNTEHHOPS---NNDLNTTK---RAERHPRKYOG 263  
 DB 815 KETAQOKPFEVHKKMEFSDAFSTKFEALQNSMOQYNOEGDAIEKHKQNSKEEYFKN 874  
 QY 264 SSY---SNLHVPCCTNTHASL----- 283  
 DB 875 ESVEEDLSRETEEOEYTHKKNFSRRKCEISAETNMREVINKIESQINYYGVIEKYS 934  
 QY 284 ---OHENSLLTK-----DRMNEKAEFCNKS-----KOPGLARSGOHR 320  
 DB 935 LIDQDQNEVSTARKLKEKIVSDLSRDKIDQYETFEFKRSTSAVENTVSTIOSLSAIDSLKR 994  
 QY 321 WAGSKETCNDRTPTTEKKVLDNADPLCERKEWKKLP-----CSENPRTDEDVPIWTL 375

DB 995 LNCISNNCKKYNT-----DIDLRSKIKTLREBQKEMPRGDKCEN---TALLKLKSL 1046  
 QY 376 NSSIQVNEWFSNDELGSDSHDGE-----SESAKV-----ADVLVNEVEDEYSG 424  
 DB 1047 RDMRGKINE---KLNDRLNSLDTKREDLLKFYSESKSHLSKDKQKPOPLNRIDEMED 1104  
 QY 425 SSEKIDLL-----ASDPHEALCKSEPRVSKSVESNIEDKIGKGYRRKASLPNLSHY 478  
 DB 1105 IKRDVDELNVNVOYISENKVTLFKNNNSVYIEMHSHINTVANGITSNNKELISKVEYE 1164  
 QY 479 ENLITAFVTEPOIIOERPLTNLKRKRRTSGIAPREDFIKADLAVOQTPEMINQTNQ 538  
 DB 1165 DKL-----NLVQNEDEYKVKKNPEN-----EKQLEAIRGSMKLEVIYNNHVS 1208  
 QY 539 TEONGOVNMTTNG-----HE-----NKTGD-----STONEKNPPIESLE--K 576  
 DB 1209 MTQLESTANTLKRNAGKENEHLELNKTKRGQMDIYEKLLIAELEKGVNELKDN 1268  
 QY 577 ESAFKTKAEPISSISNMELELNHNSKAPK-----KRLRR 613  
 DB 1269 EKANKVEPPERNITGHVLEITVEKDQKAGVYEBNLSLKTIEKLIQETSDSOHELYT 1328  
 QY 614 KSTRIH---ALELVSRNLSPPNCTEIQIDSCSSSEETK-----KKRYNQPVRH 662  
 DB 1329 TSTIKHLENAGKYEYDIKRN-----BEDSIQLRKAKSLFTLDEMKEVLQOV--- 1375  
 QY 663 SRLOLMECKEPRTGAKKSNKPN-----QSKRHDSPTPELKLNTA 705  
 DB 1376 --NMNLOSAIQNAGISK--ELNELQGIYELLISTNYSTILEVKNKSSSVRFSQLAN- 1430  
 QY 706 PGSFCTCSNTSELKEFVNPSLPREEKEKLETYKVSNNADPMDLMGSRVLOTGERSVE 765  
 DB 1431 -GEFTTAEG-----EKKNASARLAELAKLEQIVKRLDYSDIDDKYKKIIEGIRRELLK 1484  
 QY 766 SSSISLVPGTGYTOESISILEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCKDNDRD 825  
 DB 1485 ESALTWESEERKFKQCCSSHMEKAGKKIEYLK-----NN 1521  
 QY 826 TEGFKYPLGHEVHNSHETSIEME-----SELDAQYLONTFKYKSKOSFALNSPQNAE 880  
 DB 1522 GDDGK-----ANITDSQMEVGNVYSKAAHHTVEAOYDKTKARC-----E 1563  
 QY 881 ECATFSAHSGSLKKOS--PRVPECEQKEENQCKNSNIRKPYGVITAGFPVVGKDXP 938  
 DB 1564 SIYAIVTKMDNLFNESLMEKVEYKCKEKKNDKAKYIAKLKP----- 1604  
 QY 939 VDNKCSIRGSRFCSSQFRGNETGLITPNKHGLLQNPYRIPLPIKSFVTKCKKNL 998  
 DB 1605 -----YDGRIKARVSENERKI-----SELKEKAKVE- 1630  
 QY 999 LLENFEHMSPPRENGENINIPSTVSTISNNIRRENVFKEASSNINEGSSINEGSSI 1058  
 DB 1631 -----KRESSQINDVSTKSLQIDNCRQOL--DSVLSNIGRAKQ-----NAL 1670  
 QY 1059 NEIGSDENIQ-----AELGRNRG--PKLNAMLRGLVQLOPEVYKOSLPSCNCKHPEIKOE 1112  
 DB 1671 QYDSDADKMSKSVLPISSELGAEKSLDKVKA-----AKKS 1704  
 QY 1113 YEEVQVNTDPSPLYISDNLEQPMGS--SHASQVCEPDDLL-----DQG---EIKED 1162  
 DB 1705 YERNLETVONEMSGRI---NVEE--GSLTIDIDKINDIENDLLKMKQYEEGLQIKREN 1758  
 QY 1163 TSEAFEND-----IKESSAVYSKSVOK-----GELSSPSPP 1192  
 DB 1759 ADKRKSNFELVGSSEINALDLPSTSIPIKULKEEYDGTGLKNYGVKMNIEHGFYKSYN- 1817  
 QY 1193 FTHTHL-----AGQYRGAKKLESSSENLSDEE-----LPQO 1227  
 DB 1818 LIETHLSNATDVSVYTFERQSLRELAEK---EEHRLRREBEVILPLNDIKKVESLKLK 1874  
 QY 1228 HLLFGKVNNIPOSOTRHSTVATECTLSKNTENLLSLKNSLDCS---NOVILAKASOE 1283

```

Db 1875 EMKKVSAEYEGMKRHTSVSQLVODMKTIVDELKTLNDISECSVNNVSIYKVKES 1934
QY 1284 HLSEETKCSASFSSQSELEDTANTNTNQPDLFGSSQKOMBHOSQ-----GV 1333
Db 1935 KHADRRRANSMYEMVT-----LANYPLSDKAKISSGMEFNAKSNKTKLELEIFSV 1989
QY 1334 GLSDKPELVSDDEERGTGLENNQEBQSDNSNGEAAASGCESETSVSEDCSGLSOSDILT 1393
Db 1990 ISNSNELLKTKIEQSDNDYIQRERESEQLAKDATDIYVTKLNFENREKLEAKNKEEYVS 2049
QY 1394 TQQRDTMQ-----NHLIKLOEMALEAVLEBHGOPS----- 1426
Db 2050 EKVREALKRSLQVEGIRCFHFNHRLNDTELELNKRWVTTYRKKRSRESGLQEMENE 2109
QY 1427 -NSYPS-----IISDSALEDRNPPOST-----SEKAVLSQKSE-----Y 1463
Db 2110 MNTYSNITTOLEGIYVAGSEKEDIKELERSNEMERNISEKISTDSKIENKSTIDELX 2169
QY 1464 PISONPEGLSADKFEVSDSSTK-----NKEPVERSSPSKC-----PSLDDRWMHS 1512
Db 2170 KLGKNCQAHWISLISYTNMKTSKKLIMINK-----EKETKCVDYIKDSSSTDGVEYET 2225
QY 1513 CSGSLQNR-NYPSOELLKVVY-----EEQLEESGPHDLTETSLPRQD-----L 1558
Db 2226 LKGFYSGKLTFSASAEIVQMDTVSYVNFARKEKESL--NAIDIKRELTLFHQNDISIV 2283
QY 1559 EGTPLYEGISLFDPSDESDPESDRAPESAR-VGNIPSTSAKYPOLKVAASQSPA 1617
Db 2284 EGG---VQNMALY-----DKLNEKRMDELRYRISTKTKOME----- 2320
QY 1618 HTTDTAG-----YNAMESVSREKPELTASTERNKRM-----SNVSGL--TPEEFMLV 1665
Db 2321 HSTDFKPMIELHKMNETNNKSLKEKELKSVNDHMSMAEMIKNGLKTYTPESVQNI 2380
QY 1666 -----YKFAKHH-----ITLTN-----LTTEETHYV 1688
Db 2381 NNISVYIEAEVKTLEIDRDYGDNYOIVEEHKQESILDRTNALMDIEIKENNYNL 2440
QY 1689 MKTDAEFYCERLTAKFLGIAGKVV---VSFYMTQSIKERKMLNHEDEFGDVANGN 1745
Db 2441 MEVMTETI-HRYNDIEKIT-NKLYQAKTEYQIILEINIKQNDMLQNIPLKVSII--EX 2496
QY 1746 HOGKRRARES 1755
Db 2497 FENYKKKES 2506

```

## RESULT 6

```

YD86_SCHPO
ID YD86_SCHPO STANDARD: PRT: 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F3.06c in chromosome I.
GN SPAC1F3.06c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

```

```

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Roben J., Grymopiez B.,
RA Weljens I., Vansteleels E., Rieger M., Scheifer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Jelaune V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochem M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT *The genome sequence of Schizosaccharomyces pombe.*;
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: 270690; CAA94624.1; -
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3E480CA06171D9DA CRC64;

```

Query Match 3.5%; Score 339; DB 1; Length 1957;  
Best local similarity 18.5%; Pred. No. 6.7e-07;  
Matches 354; Conservative 325; Mismatches 690; Indels 548; Gaps 79;

```

QY 3 USALRVEEYQNYINA-----MOKITEPT-----CLELKEPVSTYCDHIFCFMUKL 51
Db 266 VSTLR--QTENSLRACKETLQEKLEKCAINEDSKLEELKHANYANSDAIVHDKDLIED 323
QY 52 LNOK-----KGPQCLCKNDITKR-----SLOES--TRFQOLVEELIKITICAROLD 96
Db 324 LSTRISEDNLKSEDDTLSTNKEKLEKLLNRTISGLDSKRTSNSQLEENEL----- 376
QY 97 TGLEVANSYNPAKKNNSPEH--LKDEYSIIQSMGYRRNARLRLLQSEPNPSLOETSLAV 154
Db 377 -----KESNRTIHSQLTDAESKLSFQEQEKSLKSGSIDEYQNNLSKDKMYK 423
QY 155 QLSN-----LGTVRLTKRKRIQKQTSVYELGSD-----SSEDTYVKAT 195
Db 424 QVSSQLEEARSSLAHATGKLAELINSEDFQKKIKDEKITEQDLRACINSSNELKESA 483
QY 196 YCSVGDDELLTTPGRTDEISLDSAKKACAFSEDTVDYTNHHQPSNNDLNTEKRAAE 255
Db 484 LIDKQDQELNNIREO--IKQKKVSESTOSSIQSLQDRILN-----EKR--- 525
QY 256 RHPEKYQGSVSNLHVEPCGTTHAASSLOHENSLLLTQDRMVEKAFCNKSQPGLAR 315
Db 526 KH-EVYE--SQNLELKG---LQTEISNSEHLSQL----- 555
QY 316 SOHNRMASSKETCNDKRTPTTEKKVDLNLADPLCE--KRWAKKQKLPCEPNPROTEDEVPI 373
Db 556 ---STLAKEBAVAATNNELSEKNSLSQT--LCNAPQEKRLAKSYMOLKENQON---FS 605
QY 374 TLNNSIQKVNEMFSSDELGSDSHQSESPSNKAVDVLVNL-----EVEYSGSSSE 427
Db 606 SLDTSFKKLNE---SHQEL---ENNHTITKQLKDTSSKLOQLOLERANFEQKESTLSDE 659
QY 428 KIDLA-----SDPEALICKSEVSHSKSVESNIEDKIFGTYKK-----KASLPNL 474
Db 660 NNDLFTKLLKLEESKSLIKKQEDV--DSLEKNIO--TLNEDLRKLSSEALRFLKLEKNL 715
QY 475 SHVTENLITGAFTVEPQIIOERPLTNKLRKKRPTSGLHPBDIKKADLAVQKTP--EMI 532

```

Db 716 REVIDNL---KCHETLEAORNDLHSSLDANK-TNALISELTKSEDEVKRLTANVEL 771  
 QY 533 NOGTNOTEON-----GOVANNITSGHENKTGSDSIONEKNPNPIESLEKESAFKAEPI 587  
 Db 772 TQDSKAMKOSFTSLVANSYOSISLHLYHELDHNMOSNN-----TLESESKLKTCEU 827  
 QY 588 SSS---ISNMEELINHNSKAPKRNLRKRSSTRIHALVELVSNLS-----632  
 Db 828 TQONMTLJINVOVKLMKHYNOESKVELKEVNGKLSLDLKNLSSLNVAISDNDQILJOL 887  
 QY 633 ---PNCSTLOJIDSCSSSEIKKKKKNOMPVRHSRLQ-----MEGEPATGAKSKK 683  
 Db 888 AELSKVYDSEDSOALNSGLKSLAEKO-LHTEHEELHRLDKLTGKLTIESSKSL 946  
 QY 684 PNEOTSKRHSDPELPKLTNAAGSFTKCSNTSELKEFVNPSPREKEE---KLETVK 739  
 Db 947 GKULTARQEE---ISMLKEENMSQOALTSYKSLDETLSSKLEADIEHLKKNKSEVE 1003  
 QY 740 VSNNAEDPKDMLSGERVLTQTERSVSSISLVPGTDTQESISLLEVSTLGAKTEPN 799  
 Db 1004 VERNL-----LLASNERLMD-----1018  
 QY 800 KCVSOCAAFENPKGLHGSKDNNDTEGKPYPLGHEVNHRSRTSIEMESELDAQYLO 859  
 Db 1019 -----DLKNGENIA-----SLQTELEKRAENDD--LOS 1046  
 QY 860 TFKV---SKROSFALFENPGNAEBECATFSAHSGSLKQSPKVFPECKEENCK--NNS 915  
 Db 1047 KLSVYSSEVENLLISSQTN-----KSLBKTQNKLYTEKNQKLLDK 1090  
 QY 916 NIKPVQVITAGFPVVGQDKRPVNAKCSIKGSRFCLSSQFRNGETGLTPNKHGLQ 975  
 Db 1091 DQHNVELLELTSKYKGLGEENQIKDELLALRKSK-----KQHDLOA 1133  
 QY 976 NPIRIPPLPISFVTKCKKLNLEENFEHNSMPRENGENIETVSTISNNITRENV 1035  
 Db 1134 N-----FV-----DOLKEKSDALE-QLTNEKNELVLEDSNSNNEL 1170  
 QY 1036 FKEAS---SSNINEVGSTNEVGSSINEIGSSDENIDAEGRNRPALNMLRGVLYQPEV 1093  
 Db 1171 VERISDLARLSDMKKSLSDSNVIVISDLVRVDEL-----DT 1211  
 QY 1094 YKQSLPGSNCKHPEIKQYEEVQVTV---NTDFPYLLISDMLQDPMGSSHASQVCEPT 1150  
 Db 1212 LKKDKDSLSTQYSEV-CQDRDLDSLKCESEFNKVAVS-----LRELTCKSE 1259  
 QY 1151 DD-----LLDDEIKEDTSFA-----ENDIKESSAV-FSKVQKGLSRSPPT 1194  
 Db 1260 IDVPVSEILDDNFVENAGNFSLSRLTVLSLENYLDAFNOVKEKKELDNRLLTTDAEPT 1319  
 QY 1195 -----HTHLAGYRRG--AKKLESSEENLSEDEELPCFOHLLFGKVNINPSQST 1242  
 Db 1320 KYVADLEKLOHEHDDMLIORGLERALKSEKNFLKREKM-----TENHLSLE 1369  
 QY 1243 RSHVATEC--LSKTEENLIS--LKNSLDCSNQVILAKASQEHLSSEETKCSALPS 1297  
 Db 1370 GKEETKEKIAELSSRLDNOATNKLKNQD-----HLMOEIRLKRQDVAK 1414  
 QY 1298 SQCS---ELEDTLANTQDPLIGSSKOMRHOSESQGVLSDELVSDEBERGTCLEEN 1354  
 Db 1415 EKESLILISLEESLSNOROKESSLLDKNLEH-----MLDTSRNSLSMK 1461  
 QY 1355 NOE-EOSMDSNLGEAAGSESETSVSEDCSGLSSOSDILTQOORDTMOHNLKLOEOME 1413  
 Db 1462 IBSINSLLDDKSELASAVEKGLALOK-----LHSELSLMENTKSLQDAKEKIQVDEST 1517  
 QY 1414 LEAVLEQHSQPSNST-----PSLIISD-SSALEDLRN--PEOSTSEKAVLTSOKSSEY 1463  
 Db 1518 IOE-LDHEITASKNNYEGLKNDKSIIRIDSENIQNLMLAEKSAVAKRLSTPEKSEIL 1576  
 QY 1464 PISQNEGSLADKFEVSADSSSTSKNKEPVERSPSKCSLDRWYMHSCSGSLQNRNP 1523  
 Db 1577 QFNSRLADLEHYHKSQVSESELGRSKL-----LASTTEELQLAENE 1616

QY 1524 SOBELIKVVDVEEQOLESGPHDTLETYSLLPQDLEGTPLYESGIS-----1569  
 Db 1617 RLSLTTRMLDLOQ-----VKDLSNIKDLSLSDLTLSLEBSVASLOKECKIKSNV 1670  
 QY 1570 ----LFSDDPESDPEDRAPESA-----RVGNIPSTALKVPQKVAESAOSPAAHT 1619  
 Db 1671 SLQDVLTSYQVARNAELEDEVSRSVDKIRRRDDCEHLSCGLKHLHSLQLEQHEFFRAQ 1730  
 QY 1620 TDTAGYNAMEESYREKPELTASTERVNRKMSVWSGLPDEEMLYKKF-----1668  
 Db 1731 QRRTOGLFLETKYKKO-----EKLLKLNLRQOLIPRSSILVYESYIRDEKEIIV 1782  
 QY 1669 ---ARKHHTILNLTIT-----ETHVYVAKTDAEFVCEKTLKFLGIAGKVVASY 1716  
 Db 1783 LQERLNGIELSQOLPRGICYGYFFKTRVMEVEVLDSTKQVAKIQFL--AGAETIVAF 1837

RESULT 7  
 MLPL YEAST  
 ID MLPL YEAST STANDARD; PRT: 1875 AA.  
 AC 002455;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin-like protein MLPL.  
 GN MLPL OR YKR095W OR YKR415.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C;  
 RX MEDLINE=93247549; PubMed=8483450;  
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;  
 RA "A new yeast gene with a myosin-like heptad repeat structure.";  
 RL Mol. Gen. Genet. 237:359-369(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94205265; PubMed=8154186;  
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
 RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three  
 RT new open reading frames.";  
 RL Yeast 9:1349-1354(1993).  
 CC  
 CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA  
 CC REPAIR.  
 CC  
 CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.  
 CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L01992; AAA34783.1;  
 CC DR EMBL: X73541; CAAS1948.1;  
 CC DR EMBL: Z28320; CAA82174.1;  
 CC DR PIR: S38173; S38173.  
 CC DR SGD: S0001803; MPL1.  
 CC Coiled coil; DNA repair.  
 CC KW Coiled coil; DNA repair.  
 CC FT DOMAIN 69 487  
 CC FT DOMAIN 531 1678  
 CC FT DOMAIN 1834 1866  
 CC FT CONFLICT 301 301  
 CC FT SEQUENCE 1875 AA: 218455 MW: 68340D34C906867 CRC64;  
 Query Match 3.4%; Score 327; DB 1; Length 1875;

Best Local Similarity 18.7%; Pred. No. 2e-06;  
Matches 300; Conservative 269; Mismatches 603; Indels 432; Gaps 70;

4 SALREVEQVNVINAM-----OKLEPCIELEIKPEVSTKCDHIFCKFCMLKLLNOK 55  
438 AALLLEHSHNEKNAVKELNAKNQKLEENDLQITTKORLDCROIYLLITNSVSNOS 497  
56 KGP-----SQCPLCKNDITKRSLOESTRPSOLVE-----ELKII-CA 92  
498 KGPLKEEIQIIONIOEDDSTITESDSQKVYTERVEFKNIIOLEKNAELLKAVRNLA 557  
93 FOLDGLEXYANYNNAKKNENSPHILKDVSIIOG--WGYRNRAKLLLOSEPBNLSLOET 150  
558 DKLES--KEKKSQSOLOKIESEFTVNEAKAIIITLSEKMDLESRIEE--LOKELEELKTSVP 615  
151 SLVSQSLNGVRLTKORIOPO-----KTSV 178  
616 NEDASYSVNTTKQLTETKRDLESQVODIOTRISQITRESTENMSLINKETIOLDYDSKOI 675  
179 YIELGSDS-----EDTVNKA 194  
676 SIKIGKEKSSRIAEERFKLSNTLDTLTKAENDQLRKRDYLONTILKQDSKTHETLNEY 735  
195 TYC-----SVGOELLQITPO-----GTRDEISLSQAKKAAACEFSETDVYTNETHOP 241  
736 VSCSKSLIVETELLNLKEEOKLRVHLEKNLQELNKLSPKDSLRIM---VTOLOQTOK 792  
242 SNNDNTEKRAEERHPEKYGSSVSNLHVEPGNTNTHASSLOHENSLLTKDRNVE- 300  
793 EREDLEETRKSCQCKIDEL-DAISELKKETSOQDHIKOLEEDNNS-----NIW 843  
301 ---KAEPCKN-----SKQPLASQHNRMAGSKETCDNRTPSTEKVD--LNAD 345  
844 YONKTEALKKYEESVITSVDSKQTDIEIKQYKKSLEIEEDKIRLHTYNNMDETIND 903  
346 PLCEKKEKNOKLPCE---NPRDVEDPWITLNSIQVNMWFSRSDLLG----- 394  
904 SL--RKELEKSKINLIDAYSQIKKEYKL-YETTSOSLOOTN---SKLDSFPOFTNQIKN 957  
395 -SDSHDESESNMAVDVLVLEVD-----EYSGSEKIDLASDPHEALICKSE 445  
958 LTDEKTSLEDKISLKEGMFNLNELDLQKGMKEKADFKRIRISILONNKEVEAVKSE 1017  
446 -----RHVSKSVESIEDKIFCKYRKKASL-PNLSHTENTLIIIGAFYTE 489  
1018 YESKLSKIQNDLDOQTIYANTAONNYEOEL-----QKHADVSKITSELREQHTYQGYK 1072  
490 POIIQERLTKKRRRPTGSLHPEDFIKADLAVQCTPEMINQCTOTDEQGVNMTT 549  
1073 TLNLSROOLENALENKESWSS--QKESLLEDLJLDSNRIEDLSSQNKLLYDOI-QIYTA 1130  
550 NSGHEKTKGDSION-----EKNPPIESLEKESAFKTKAEPJISSIMMEL-----N 599  
1131 DKEVNNSTNGPGLNILLITLRENDILOT-KVYTAEDAKALROKISIMDELQDARTK 1188  
600 IHNSKAPKKN---LRRKSTRIHAEIVVSRNLSPPNCTEQIDSCSSE---EIKK 652  
1189 LDNSRVEKENSHSIIQCHDDIMEKLNQNLRLRESNITLRN-ELENNNNKKKELOSELKD 1246  
653 KKYQOMPVRHSRNQIMGKEPATGAKSKNKPNEQTSKRHSDTPEPLKLTNAPSFTKC 712  
1247 LKQVAVP-----ESELTAALKYSQCKED-----ELKLAK----- 1276  
713 SNTSELKEFVNPSPREKEKEKLET---VKVSNNADEPKDMLSGERY----- 757  
1277 ---BEVHRMKKRQDDILEKHQSSDOVEKLESEIENLKELENNKRGAEAEKFNRLR 1333  
758 LQTRRSVSESSISLVPGTGYQESISLLEVSTLGAKEATPNKCVSQCAA---FENPKG 813  
1334 ROADERKTSKLS-----ODSLT-EQVNSTLDANKVLENSISEANARIEELQANK- 1382  
814 LIHCSMDNRMDTGFKYPLGHEVNHRSRETSIEMSESLDAQVLYON-----TFK-VS 864

1383 -----VAQNNOLEAIIRKLOEADAERASRELQAKLEESTTSYESTINGINEIITLKEEIE 1437  
865 KROSF--ALFSPNGAEEPCATFSAHSGSLK--OSPVTPECEQKEPNOCKNSNPKPV 920  
1438 KORQIOQLOATISANDQNSLIVE---SMKSFEDKIF---IKETQEVNKKIIEAQ 1491  
921 QTVNIYAGPVPVQOKDKPYDNAKCSIKGSRCESSQSPRGNETGLTPNKHGLQNPYRI 980  
1492 ERLNQPSNINMEIETKKMWESEHQEV-----SQKIREADEA-----LKKRIRL 1534  
981 PLPLPFSVTKCKCKNNLLENEFESHSPEREKGNENIPSTVSTISNNIR----- 1032  
1535 PTEERIKNIIEER--KKEELEKEFEET---KVEERIKSMQSGEIVVILKQLEAVQEKOK 1589  
1033 --ENVFKEASSNINVEGSSNTN-----EVGSSINGISSDENIQAEQLRNNGPK 1079  
1590 ELENEYNKKLODELVDVPHSSITSDEDEDKRLAEIESKRLKEEPFNNELQAIKKSFDDG-K 1648  
1080 LNAMLRGLVLOPEVVKOSLPGSNCKHPRIKQOEVEVQVNTDPSPYLLSDNLEOP-MG 1138  
1649 QQAMMKITLLERKLAKMESQJSE-----TQSAESPCKSVN-----NVQNPILG 1692  
1139 SSHASQVCEFPDLDLDDGEIKEDTSPFANDIKESSAVFSKSVQKGLSRSPEFTHTL 1198  
1693 IPRKIEENSNSPFPNLLSGE-----KLKLKSKSSSGGFNFT-----SPSPNKH--- 1737  
1199 AGYBRGAKKLESSEENLSEDEELPCFQHLHFGVNNIPQSS---TRHSTVATECJSKN 1255  
1738 ---LQNDNDKRESLANKNDPPLHLEPSF-----NIPASRLISSSTLSD---T 1781  
1256 TEENLSLKNLSINDCSNOVILAKASQ-----HMLSETKCSASLFFSQCSL- 1303  
1782 NDEELTSNMPAQKDSNSNRVQSEEDTEKKKEGEPVKRGALEEQTKSNKRP-DEVGELK 1840  
1304 ---EDLTANTNPQDPFLGSSQKMRHQSQSGVGLSKELYSDD 1344  
1841 NDEDTTENIN-----ESKIKITEDEE---KETDKVND 1872

RESULT 8  
BRC2\_HUMAN  
ID BRC2\_HUMAN STANDARD. PRT. 3418 AA.  
AC P51587; Q13879; O00183; O15008;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Breast cancer type 2 susceptibility protein.  
GN BRC2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96112016; PubMed=8524414;  
RA Wooster R., Bignelli G., Lancaster J., Swift S., Seal S., Mangion J.,  
RA Collins N., Gregory S., Gumbs C., Micklem G., Barfoot R., Hamoudi R.,  
RA Patel S., Rice C., Biggs P., Hashim Y., Smith A., Connor F.,  
RA Arson A., Gudmundsson J., Filene D., Kelsell D., Ford D., Tonin P.,  
RA Bishop D.T., Spurr N.K., Fonder B.A.J., Beles R., Peto J., Devilee P.,  
RA Coriell C., Lynch H., Narod S., Lenoir G., Egilsson V.,  
RA Barkadottir R.B., Easton D.F., Bentley D.R., Futreal P.A.,  
RA Ashworth A., Stratton M.R.;  
RT Identification of the breast cancer susceptibility gene BRC2.  
RL Nature 378:789-792(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96172838; PubMed=8589730;  
RA Tavtigian S.V., Simard J., Rommens J., Couch F., Shattuck-Eidens D.,  
RA Neuhansen S., Meraayer S., Thorlacius S., Offit K., Stoppa-Lyonnet D.,  
RA Belanger C., Bell R., Berry S., Bogden R., Chen Q., Davis T.,  
RA Dumont M., Frye C., Hattier T., Jamulapati S., Janacki T., Jiang P.,  
RA Kener R., Leblanc J.-F., Mitchell J.T., McArthur-Morrison J.,

RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,  
 RA Stringfellow M., Strop C., Swedlund B., Swensen J., Teng D.,  
 RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,  
 RA Wong A.K.C., Shizuya H., Eyford J.E., Cannon-Albright L., Labrie F.,  
 RA Skolnick M.H., Weber B., Kamb A., Goldar D.E.,  
 RT "The complete BRCA2 gene and mutations in chromosome 13q-linked  
 RT kindreds.";  
 RL Nat. Genet. 12:333-337(1996).  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Hunt S., Murray A., Williamson H.,  
 RL submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP VARIANT OC HIS-2787, AND VARIANTS ASN-372; MET-1915 AND VAL-2466.  
 RX MEDLINE=96275740; PubMed=665505;  
 RA Takahashi H., Chiu H.-C., Bandera C.A., Behrhardt K., Liu P.C.,  
 RA Couch F.J., Weber B.L., Livolsi V.A., Furusato M., Rebani B.A.,  
 RA Cardonick A., Benjamin I., Morgan M.A., King S.A., Mikuta J.J.,  
 RA Rubin S.C., Boyd J.,  
 RT "Mutations of the BRCA2 gene in ovarian carcinomas.";  
 RL Cancer Res. 56:2738-2741(1996).  
 RN  
 RP VARIANTS ASN-372; ASP-991; SER-1147; MET-1915 AND CYS-2034.  
 RX MEDLINE=96241588; PubMed=6673091;  
 RA Couch F.J., Farid L.M., Deshano M.T., Tavtigian S.V., Calzone K.,  
 RA Campeau L., Peng Y., Bogden B., Chen O., Neuhausen S.,  
 RA Shattuck-Eidens D., Godwin A.K., Daly M., Radford D.M., Sedlacek S.,  
 RA Rommens J., Simard J., Garber J., Merajver S., Weber B.L.,  
 RT "BRCA2 germline mutations in male breast cancer cases and breast  
 RT cancer families.";  
 RL Nat. Genet. 13:123-125(1996).  
 RN  
 RP VARIANT GLU-3095.  
 RX MEDLINE=96225456; PubMed=6640235;  
 RA Lancaster J.M., Wooster R., Mangion J., Phelan C.M., Cochran C.,  
 RA Gumbs C., Seal S., Barfoot R., Collins N., Bignell G., Patel S.,  
 RA Hamoudi R., Lattson C., Wiseman R.W., Berchuck A., Iglehart J.D.,  
 RA Marks J.R., Ashworth A., Stratton M.R., Futreal P.A.,  
 RT "BRCA2 mutations in primary breast and ovarian cancers.";  
 RL Nat. Genet. 13:238-240(1996).  
 RN  
 RP VARIANTS  
 RX MEDLINE=96225457; PubMed=6640236;  
 RA Teng D.H.-F., Bogden R., Mitchell J., Baumgard M., Bell R., Berry S.,  
 RA Davis T., Ha P.C., Kehrner R., Jammulapati S., Chen O., Offit K.,  
 RA Skolnick M.H., Tavtigian S.V., Jhanwar S., Swedlund B., Wong A.K.C.,  
 RA Kamb A.,  
 RT "Low incidence of BRCA2 mutations in breast carcinoma and other  
 RT cancers.";  
 RL Nat. Genet. 13:241-244(1996).  
 RN  
 RP VARIANT ASN-2415.  
 RX MEDLINE=96225458; PubMed=6640237;  
 RA Miki Y., Katagiri T., Kasumi F., Yoshimoto T., Nakamura Y.,  
 RT "Mutation analysis in the BRCA2 gene in primary breast cancers.";  
 RL Nat. Genet. 13:245-247(1996).  
 RN  
 RP VARIANT BC ASP-2089, AND VARIANT VAL-3412.  
 RX MEDLINE=97294396; PubMed=9150152;  
 RA Veemanen P., Friedeman L.S., Eerola H., Sarantausta L., Pylhonen S.,  
 RA Ponder B.A.J., Muhoonen T., Nevalinna H.,  
 RT "A low proportion of BRCA2 mutations in Finnish breast cancer  
 RT families.";  
 RL Am. J. Hum. Genet. 60:1050-1058(1997).  
 RN  
 RP VARIANT BC AND PANCREAS CANCER TRP-554.  
 RX MEDLINE=98316775; PubMed=9654203;  
 RA Ganguly T., Dhulipala R., Goddard L., Ganguly A.,  
 RT "High throughput fluorescence-based conformation-sensitive gel  
 RT electrophoresis (F-CSE) identifies six unique BRCA2 mutations and an  
 RT overall low incidence of BRCA2 mutations in high-risk BRCA1-negative  
 RT breast cancer families.";  
 RL Hum. Genet. 102:549-556(1998).

RN [11]  
 RP VARIANTS BC L-32; R-53; L-81; R-201; A-211; S-222 AND T-3118.  
 RX MEDLINE=98272917; PubMed=9609997;  
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,  
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Naha K.,  
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furusuma Y.,  
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,  
 RA Hougou S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,  
 RA Sonoo H., Kurebayashi J.-T., Shimotsu K., Nakamura Y., Miki Y.,  
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in  
 RT Japanese breast cancer families.";  
 RL J. Hum. Genet. 43:42-48(1998).  
 RN  
 RP VARIANTS OC PRO-75; HIS-2502 AND HIS-3098.  
 RX MEDLINE=20455732; PubMed=10486320;  
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,  
 RA Ponder B.A.J.,  
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial  
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility  
 RT genes.";  
 RL Am. J. Hum. Genet. 65:1021-1029(1999).  
 RN  
 RP VARIANTS HIS-289; ASN-372; ASP-991 AND VAL-3412.  
 RX MEDLINE=99254821; PubMed=10323242;  
 RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,  
 RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.-H., Tseng J.-N., Chen A.,  
 RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.,  
 RT "Molecular characterization of germline mutations in the BRCA1 and  
 RT BRCA2 genes from breast cancer families in Taiwan.";  
 RL Hum. Genet. 104:201-204(1999).  
 RN  
 RP VARIANTS BC, AND VARIANTS.  
 RX MEDLINE=99138688; PubMed=9971877;  
 RA Wagner T.M.U., Hirtelshner K., Shen P., Moeslinger R., Muhr D.,  
 RA Fleischmann E., Concin H., Doeller W., Haid A., Lang A.H., Mayer P.,  
 RA Petru E., Ropp E., Langbauer G., Kubista E., Scheiner O.,  
 RA Underhill P., Mountain J., Stierer M., Zielinski C., Oefner P.,  
 RT "Global sequence diversity of BRCA2: analysis of 71 breast cancer  
 RT families and 95 control individuals of worldwide populations.";  
 RL Hum. Mol. Genet. 8:413-423(1999).  
 CC  
 CC -I- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE  
 CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS  
 CC RECOMBINATION.  
 CC  
 CC -I- SUBUNIT: INTERACTS WITH RAD51.  
 CC  
 CC -I- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION IN BREAST AND  
 CC THYMUS, WITH SLIGHTLY LOWER LEVELS IN LUNG, OVARY, AND SPLEEN.  
 CC  
 CC -I- DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,  
 CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE  
 CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF  
 CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-  
 CC ONSET BREAST CANCER. MUTATIONS IN BRCA2 ARE THOUGHT TO BE  
 CC RESPONSIBLE FOR SOME INHERITED BREAST CANCER. IT IS LINKED WITH  
 CC MALE BREAST CANCER.  
 CC  
 CC -I- SIMILARITY: CONTAINS 8 BRCA2 REPEATS.  
 CC  
 CC -----  
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION  
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X95152; CA64484.1; JOINED.  
 DR EMBL: X95153; CA64484.1; JOINED.  
 DR EMBL: X95154; CA64484.1; JOINED.  
 DR EMBL: X95155; CA64484.1; JOINED.  
 DR EMBL: X95156; CA64484.1; JOINED.  
 DR EMBL: X95157; CA64484.1; JOINED.  
 DR EMBL: X95158; CA64484.1; JOINED.  
 DR EMBL: X95159; CA64484.1; JOINED.  
 DR EMBL: X95160; CA64484.1; JOINED.  
 DR EMBL: X95161; CA64484.1; JOINED.









```

QY 703 TNAPGSPKCS--NTSELKEFPNPSLPREKEEKLFTKVSNNAEDEKDLMLSGERYLOT 760
D 2107 KVP-PROPISEHOTREVEOLAN---HLKRTDKCSLLLSKE-QLOORDQENRELEKL 2161
QY 761 ERSVESSISLVPTDYGTQOESISLLEVLTLGAKAKEPNKCVCOCAPFENPKGLIHGCSK 820
D 2162 EFRRELEOALIVSAD--TFQKVE--DRKHFGAVEARPE--LSLEVOLOQAEPAIDRREK 2215
QY 821 DNRDTEGFPYPLGHEVNSHRETSIEMESELDAQYQNTFNFKVSKRSPFLFSPGNAE 880
D 2216 EITN-----LEBOLEQFRE--ELEKNNEEVOQLHMOLELOKKESTTRLO--ELEQ 2261
QY 881 ECATFS-----AHSGSLKQSPKVTPE-----CEQKE-ENQCKNSNIIKPVQTVN 924
D 2262 ENKLFKDMKELGLAITKESDAMSTQDOHVLFGRKFAQIIOKEVEYIDOLNQVYKRLQOOLK 2321
QY 925 ITAGFPVVGOKDPVDNAKCSIKGSRCLSSQFRGNETGLITPNKHGLQNPYRIPPLF 984
D 2322 ITTDNKVIEEKNELIRDLFOIE-----CLMSD----- 2349
QY 985 PISFVTKCKKNLLENEEE-----HMSPER----- 1012
D 2350 -----QECVKRREBEIOLNEVIEKLOQELANIGOKTSMANHSLEADSLKHQLDV 2402
QY 1013 -----EMGNENIPSTVSTISRNIRENVK-----EASSNINEVGS 1049
D 2403 VIAEKLEQOQVETANEEM-----TEPKNVKLETFNFMNOLQELFSLKRESRVEKIOS 2457
QY 1050 -STINEVGSSINELGSSDENIOAELGRN-----RGPKINAMLRGLVQ 1090
D 2458 IPNSVVAADHLSKDKPELEVVLTEBDALSKLENQYFKSFEENGKSIINLETRLOLE 2517
QY 1091 PEVYKOSLPQSNCKHPKPEIKKOEVEYQVNTDFSPYLIDNLEQPGSSHAQVOCSETP 1150
D 2518 STYSAKDLDTQC-KYOKIMQOGOFETMLOKKIYNLCKIYEKVAALVSOIQLEA- 2575
QY 1151 DDLIDGEEKEDTSFAENDIKESSAVFSKSVQ-----KGLSRSPPTF----- 1194
D 2576 -----VOEYAKFCODNOTISSEPERTNIQNLNOLREDELSDSLALRISLESOY 2627
QY 1195 ---HTHLAGYRGAKAKLESSENLSSEDELPFOHLLFGKYNINIPSGSTRHSTVATEC 1251
D 2628 VEMHTSLI-----LEKQVEIAEKNVLEKEKLELOLLEGENEKKOREKREKSSPODEV 2683
QY 1252 LSKNTEENLSTKNSLNDSCNOYLAKASQEHHLSEPTKSASLFSQSCSELDLTANTN 1311
D 2684 LKTTTTE---LPHSNESGFFNELALRAE-----SVATKALASYPEKAEKIELOEELLVKE 2735
QY 1312 TODPFLIGSSKOMR-HQSES-QGVGLSDKELVSDDEER----- 1347
D 2736 TMTSTLOKDLQSYRDLHALEAKELSTLEKDETEVEQESKACMFEPPLIKLSIASQTD 2795
QY 1348 GT-GLENNQOEDOSMDSNLG---EASGCESETSVSEDCGSLSSQ-----SDLTLT 1394
D 2796 GTLKISSNOTPOLVKMGIQINLOESGSE-EVEIITISOFTKEIEKMOELHAAETLDM 2854
QY 1395 QOBDTMOHNLK-----LQOEMAELEAVLE----- 1419
D 2855 ESNHISETEFLKREHYVAVOLLEBEGGTAKAVIOLCLRSKEVFGFYNNCSTLDCDSGDMG 2914
QY 1420 -----QHGSQPSNSYP----- 1430
D 2915 OGILYTHSQOFLIASGREGSESATDSPFKIKGLLRVHNEGMOVLSTESPYDGED 2974
QY 1431 -----STISSSALIEDLRNEQSTSEKAVLTSCSSS----- 1462
D 2975 HSIQOVSPEPWELEBKRYININTISLKLITKMOQRAEYVDSQSSESPDMRGELLAL 3034
QY 1463 --YPISONPEGLSADKFEVADSST-----SKNKEPGEVRSRSPSCPLDPRWM 1510
D 3035 QOVFLERSVLLAARFELTALGTTDVAVGLNCLBQRIQOGVEYQVQAMECLOKADR--- 3091

```

```

QY 1511 HSCSGSLQNNRNPYSQEBELI-----KVVDVEEOQLESQGHDLTETSYLPRODEGTPYL 1564
D 3092 -----KSLSEIALHQAQNNGRKITLKEDESKPSELLEYNIQOQ-----SOML 3138
QY 1565 ESGISLSDPESDPSDRAPESARVGNIPSSISALVPOLKAESQSPAANTPTAG 1624
D 3139 EMQVEILSS-----MKDRATELOE-----OLSSKKVVAEIK-SELQOTYLETLETTLKAQ 3186
QY 1625 YNMBE-----SVSRKEPDLASTPERVAKRM 1650
D 3187 HKHLKELEARLEKVKQDTEVHLINDTLASEQKSKRELQWALEREKAKLGRSEERKEEL 3246
QY 1651 SMVSGLTPEEFMVLVYFKARKHHTLNLTEFTTHVAKMTDAFVCERTL 1701
D 3247 -----EDLKFSLESQKQNRNQL-NLLLEQKQOLLNSQOKIESQRM 3287

RESULT 10
CENT HUMAN
ID CENT HUMAN STANDARD: PRT; 3210 AA.
AC PA9454: Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (centromere protein F) (Mitosis) (AH antigen).
GN CENP-F.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eultheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95379848; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=9536446; PubMed=7612011;
RA Li O., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC - FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
CC - SUBUNIT: HOMO- OR HETERO-DIMER.
CC - SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),

```

```

CC REORGANIZATION TO THE KINETOCORE/CENTROMERE (CORONAL SURFACE OF
CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
CC DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
CC -1- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL; U19769; AAA82889.1; -
CC EMBL; U30872; AAA82935.1; -
CC EMBL; U25725; AAA86889.1; -
CC Genew; HGNC:1857; CENPF.
CC MIM: 600236; -
CC -----
DR InterPro; IPR001230; Prenyl_site.
DR Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT COILED COIL (POTENTIAL).
FT DOMAIN 273 769
FT COILED COIL (POTENTIAL).
FT DOMAIN 823 1328
FT COILED COIL (POTENTIAL).
FT DOMAIN 1642 1746
FT COILED COIL (POTENTIAL).
FT DOMAIN 1862 2568
FT COILED COIL (POTENTIAL).
FT REPEAT 2207 2386
FT 2 x 177 AA TANDEM REPEATS.
FT REPEAT 2389 2568
FT 2.
FT DOMAIN 3015 3032
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 16 16
FT T -> A (IN REF. 2).
FT CONFLICT 250 250
FT L -> Q (IN REF. 2).
FT CONFLICT 272 272
FT G -> D (IN REF. 2).
FT CONFLICT 611 611
FT MISSING (IN REF. 2).
FT CONFLICT 1494 1589
FT MISSING (IN REF. 2).
FT CONFLICT 1611 1611
FT V -> A (IN REF. 2).
FT CONFLICT 1811 1811
FT V -> L (IN REF. 2).
FT CONFLICT 2242 2243
FT ER -> DG (IN REF. 3).
FT CONFLICT 2335 2335
FT L -> Q (IN REF. 3).
FT CONFLICT 2492 2492
FT D -> N (IN REF. 3).
FT CONFLICT 2545 2561
FT ELNERYAALHNDQEQACK -> SSMREMOQCMTKKRVVS
FT (IN REF. 3)
SO SEQUENCE 3210 AA: 367589 MW: 110832496064334 CRC64;

Query Match 3.4%; Score 323.5; DB 1; Length 3210;
Best Local Similarity 19.2%; Pred. No. 5.6e-06;
Matches 392; Conservative 290; Mismatches 735; Indels 621; Gaps 93;

QY 9 EGVGVVIAAMQ-----KILEPCITL-----LIKEPVSTKC----- 39
DB 288 QELRNKINLELELRLOQHEKEMKGOVNRKFQELQLOLEKAVELIEKEKVLNKCDELVRTT 347
QY 40 ---DIIFGFCMLKLNOKKGPSCPLCKNDITKRSLOSTRPS-----OLVEEL 86
DB 348 ALOYDASRTYTLAEQKLK-----LTEDLSCQRONAESARCSLEQKIKEKEKEOEEL 400
QY 87 LKIICAFQ-----LDTGLEYA-NSYNFAKKE-----NNSPEHLKD-- 120
DB 401 SROGSPQTLDECIOMKARLTQELQOAKNMHNVLOALEDKLTSYKQOLNNLEEFKQKL 460
QY 121 -----EYTIOSMGYRNRAKRLLOSEPENPSLOETSLSYOVSINIGYVTLRT 167
DB 461 CRAEQAFOASQIKENELRSMEMKRNENLLSHSQOKAREVCHLEAELEKNI----- 512
QY 168 KORIOPKTSYIETGSDSESDTYNKATYCSVGDDELLQITPGTRDEISLSAKKACE 227
DB 513 KQCLAQSONFAEMAKNTSOETMLRDLOEKINOE-----NSLTLEKLAVAD 562
QY 228 F-----SETDVNTENH-QPSNNDLNTTEKR-----AABRHPEKYOGSSVSNLHVEPC 274
DB 563 LEKQDCSQDILKREHNIEQLNDKLTSEKESKALLSALDELKKYE-----ELK 613
QY 275 GTNHASSLOHENSLL--LTQDRANVEK-----AEPCNKSQOPGLARSOHNMAQSKETFC 328

```

```

DB 614 EEKTLFSCWKSSEKLLTOMESEKENLOSKINHLETCCTKTOO--IKSHENYERVATLTM- 670
QY 329 NDRRPTSTE-----KQVDLADPLC-----ERKEMNOK-----LPCSENP 364
DB 671 -DRENLSVEIRLNHVLNDSKSVETLOKLAIVLEQAKFAFSQOKOKELENNCKLTSTLT 729
QY 365 RDEDVPIWTLNSSIOKVNEFWSRSDLLGSDSHDESESNKAVADLVLENYDE--Y 422
DB 730 GQVED-----LEKIQDLT-----SNEIMDKRCY--QDLAAEVESLRDLKSKDASLV 775
QY 423 SGSESEKIDLLASPHALICKSERVNSKVESNIEDKIFGKYRKKAISPLNHWTEML 482
DB 776 TNEHQRSLLAFDQOPAM-----HHS-----FANIGQGSMPs----- 809
QY 483 ICAFTVPEQIOERPLTNKTKRRRRTSGLHDEDFIKKADLVAQKTPMINQGTQTEQN 542
DB 810 -----ERS-ECRLDQSPKNAIIQNRVDSLEFSLSEQOKMNSDLQKQCE-- 854
QY 543 GOVMNITNSGHENKTKGDSIONE--KNPNPLESLEK-SAEKTAEDPISSTISNMELEL 598
DB 855 -ELVQKGEIEENLMAEQMHQSFAETQSQISKLOEDPSAHQNVVAFTLSALEKKEKL 913
QY 599 NIHNSKAPKKNRLKRSSTRHIALELVYSRNLSPNCTELOIDSCSSSEETKKKKNQM 658
DB 914 QOLNDRV-----ETEOAEIOELKSNHLLDPSKELQTL--LSETLISLEKEMSSI 961
QY 659 PVHRSNLO-----LME-----GKEPATGAKSNPNEDTSKRHD-----SDTFPEL 700
DB 962 ISLNKREIELTQENCTLKEINASLNQEKMLNLOKSESPANYIDEREKISIELSDQKOE 1021
QY 701 KLTNAPGFTKCSNTSELKEFVNPSTL-PREEKEKLETV----- 738
DB 1022 KLI-----LLQREPEFGNAVEDLSQYKKAQEKNSLLECLINCTSCENRKNLEQLKEA 1077
QY 739 -----KYSNNAEDPKDMLSGERYLOTERTV-----ESSISLVPGTDTQSTIS 784
DB 1078 FAKHOEFLTKLAFEBERNQNMLETETVOQLRSEMTDNONNSKSEAGGL--KOELMT 1134
QY 785 LLEVSTLGAKTEPKKVCVSCAFAENPKGLJGC-----SKDRNRTGEFKYPLGHEVNS 840
DB 1135 LKEEQ--NMQKEVNDLLOENQOLKVKYKTKKECNLSSEPTKRSVKE--RESERQCNFK 1191
QY 841 RETSIEMESSELDQYLONTFRVSKROSFALFNSPGNAEEECATFSAHSGLSKQSPKVT 900
DB 1192 PQMDLEVKREISD-----SYNAQLVQLEAMLRNK-----ELKIQESPK 1230
QY 901 FPCQKEENQGNKNE--SNIRKYQVNTIAGRPVVGQDKRV----- 939
DB 1231 KECLOHELOTIRGDLTSLNLOMOQOELS-----GLKDCEIDAEKYSIGPHELSTSON 1284
QY 940 DNAK-----CSIKGSRFLSSQPGNTEGLTTPK----- 970
DB 1285 DNANHLOCSLOTTMKNKLNLEKICELLOAEKVELYELNDSSECTTARKMAEYGLLN 1344
QY 971 -----HGLL-----ONPRIPRLPEPISFYK--TKCKRNL----- 998
DB 1345 EVKILNDSGLLHGELVEDIPGEGEOPRNBQHPVSLARLDESNYSYELLTLSDREYOMHF 1404
QY 999 --LEENF--EHSMSPER-----ENGNNISTYSTISRN----- 1029
DB 1405 AELQEFKLSLOSEHRTILDONCOMSKMSSELQTYVDSLAKAENL--VLSTNLRNQGDLVK 1462
QY 1030 -----NIREVFEKAESSNINEVSGSTNFGSS-----INFGSSD--ENIOALGLNRG 1077
DB 1463 EMQOLGEBGLVPSLSSCVYP--SSLSLGSBSPFRALLBETGDMSLNIEGAVSANQ- 1520
QY 1078 PRLNMLRLGLVPEVYKOSLPGSNCKHPEIKQOEYEVQTVN--TDFSPYLLSD--NL 1133
DB 1521 -----CSVDVFCSSL-----QTYVDSIKAEKLVLTNLRNQGDLVKEMQGL 1564
QY 1134 EQPGSSHAQVOCSETPDDLDLDDGELIKEDTFAENDIKES----- 1173

```

```

Db 1565 EEEGLVPSLSSS-CVDPSSSLSSLG---DSSFYRALLEQTDGMSLLNLEGVVSANQCSV 1619
OY 1174 SAVFSKSVQGGELSRSPSPFTHT-----HLAOGVYRGAKKLESS----- 1212
Db 1620 DEVFCSLSOEENLTRKTPSPAKVGEELSLCEVYKQSLKELEKMEGOSIMKKEITOE 1679
OY 1213 -EENLSEDEELPCFQHLFGKYNIPQST-----RHSTVATECLSKNEENLLSKNS.1266
Db 1680 LEQULSSERELDCLRKQYLSBNQMOQKLTSLVLEHESKLAEE--KKQEQSLSELEVA 1737
OY 1267 L-----NDCNOYILAKAQQEHLSEETKCSASLFPSSQCELEDLTANTNTQDPFLIGSS 1321
Db 1738 RLQLOGJLSSRSRLIGDTEDAIOGRNESCDS-----KEHTSETTERP----- 1782
OY 1322 KQMRHOSQSGVGLSPELSD--DEER--GTGLENNQ-----EQGMDNL---GB--- 1367
Db 1783 KHHVHQ-----ICDKAQODLNLDEKLTTEGAVKPTGECGSEOSPDNTYPPGDKT 1835
OY 1368 -AASGESETSVS-----EDCSGLSSQSDILTTQOR--DTMOHNL-----IK 1406
Db 1836 QGSSECTSELSEFGPNALVPMDFLGNQEDIHNLQLRKETSNNRLHLHYIEDRKRYES 1695
OY 1407 LQDPMALAVLELQHSQSPNSIPIISDSSALEDLRNPEOSTSEKAVLTSQSSSE-YPI 1465
Db 1896 LLEMEELDSKHLQEVQYLTMTKLEACIELEKIVGELKENSDELSEKLEFSCDHOELLQ 1955
OY 1466 SONPEGLSADKPFVSADSTSKNKEPVERSSPKCPSLDRLWYMHSCSGSLQNRNYPQ 1525
Db 1956 VETSEGLNSD-LEMHADKSSREDIGDNVAK-----VDSW-----KERFLDVE 1997
OY 1526 EELIKV---VDVEEQULESGPHDLTETSYPRLDLEGTPTLESGLSPDSDPESPSE 1581
Db 1998 NELSRIRSEKASILEHLYLEADLEVYQTEKL---CLEKDNENKQKQYVLEELSVYTS 2054
OY 1582 DRAPESKRVNINSSSALKAVPOLKVAESQSPAAHTTPTAGYNAMESSVSEKPEL 1639
Db 2055 ERNQLRGELDTMSKRTYALDQLESEKMEKETOE--LESHQSECLHCIOVAEALEVEKETEL 2111

RESULT 11
YME67_YEAST STANDARD: PRT; 1658 AA.
AC 003661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Hypothetical 187.1 kDa protein in GOAL-Erg8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=5288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=5288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; Z49809; CAA89934.1; -

```

```

DR EMBL; Z49939; CAA90190.1; -
DR SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match
Best local similarity 19.7%; Pred. No. 2.8e-06;
Matches 344; Conservative 255; Mismatches 704; Indels 444; Gaps 76;

3.3%; Score 322; DB 1; Length 1658;

OY 132 NNAKRLQSEPNLSLOETLSVQLNGLTVRLTKRQRIQPOKTSYVIELGSDSDTV 191
Db 97 NRGKSLITLLEKEDALFERSLIEE-----RQRO-----LHDSLM 131
OY 192 NKAATYCVGQDELQI--TPQGT-----RDEISLDAKKAACEFSDVTNTHHQPSN 243
Db 132 NKTGNSKSHQRLDLKSOYGTDTSQNNDELPLDS-----FISSPLPADESSN 184
OY 244 NDINTTEKRAAEHRPEKYQGSVSNLHVEPCGNTTHASLQHENS-----SLLTQR- 296
Db 185 IDSQKDEDLGKQ-----SLIKDFLE--NDEYELSEBKNSDQSSPSIMILSDEE 234
OY 297 -----ANVEKAFCKSKOPGLARSOHN-----RMGSKETCDRRTPETEK 339
Db 235 YAEAGALQDVSNDEYAEEOVERKNIGQOAVENATQISSDSEGOVSEGVEMELE 294
OY 340 VDLNADPLCEKE-----NNKOKLPCSENPRTQEDVPMI--TLNSSLQKVMWF 386
Db 295 DDIDVESDAKDESQAGEGTEHVSVDPSKYMQPRTDNTK---IPVIEKYSDEKHQORY 350
OY 387 SR-----SDELLGSDSHGSESNAKVADYLDVLEVEYSGSSEKI--DLASDPHE 438
Db 351 SEDGAFDGSVNIYVDESEDEESQAESYSAKMEVYVHNHHELDKELLEDIESD--- 407
OY 439 ALICKSERVSKSVESNIEKIRGKYRKASLPNLSHTVENLIGAFVETPQIIQERPL 498
Db 408 -----SESGAQSQSEQSEDDEFKMKNEKSTSEETNTSES-----RDQGAQDAYT 455
OY 499 TNKLKRRRPTSGHLRFDIKKADLAQKTPPMINQGTNOEDQGVNMTNHSHEKTK 558
Db 456 KMKVEQDEND--EPK---KODILRSSLDKNFNGNNKSEYSENVL-----ENETD 501
OY 559 GDSIQNEKNPNPIESLEKESAFKTKAPISSISNMLELNH----- 601
Db 502 PALVERNQINDVQGV-----VTGKVEEDLHHPSPDNLYDLAARMQ 546
OY 602 -----NSAKPKKRLRRKSTRHILALELVSRNLSPPNCTELQIDSCSSSEITKKKYN 656
Db 547 FOOSRNSNCPQKE-----EQVSESYLGHSNGLSGRLDSESE----- 585
OY 657 QMEVVR-----HSRNLQMEKRPATGAKKSNKPNQTSKRHDSDTPPELKLITNAPGFTK 711
Db 586 QIPKPDGTGNNNNKLTDRGDLSSVIEYKVE--KKLDGSTEKL-----VPLSDT 638
OY 712 CSNTSEL--KEFNPSLIPREK--EEKLETYKVSNNADPMDLMSGRVLOTERSVSS 768
Db 639 TINNSSLGNDSTIYSLDADALSENLTDPIMEIKTTPYEVYISSV--YSSVSDNT 697
OY 769 ISLVP-----GTDYGTQESISILEVSTLCKATPEPKVCVQAAPENPK 812
Db 698 VAMPPOVEYTPPMNDPFNSLNDYERKHDLLKSTIALALPAETKQD-----AEFVE-A 750
OY 813 GLIHGSKQRNDEGKRYLGLGHEVNSRETSTIEMESELDQY--LQNTFKVSKROS--F 869
Db 751 GVTKSLTTSYGHNTIF-----HTSKETKQVS--DLDESTVYVFENNTDEKNOSKNF 804
OY 870 ALFSNPGMADEECATSAHSGSLKQSPKVTPECEQKEENQAGNESNI-----RPVQTVNI 925
Db 805 PGVANSKSTEDNTDEKYSALNTYNVTGSSCEDIIETASVWEELRRCEDMDMAEM 864
OY 926 TACFPVYGQDKRVYDNACKSITKGSRPCUSS--QFR-GNETGLITPKHGLLONPYRIP 982
Db 865 SSGDECVKQND--DSKTOIS---FSTDSPNFQESNDNTFFSSTKYKV-----RNSD 912

```

QY	983	LFPIKSFVKTCKNKL-----EENPHEHMSPE--REMGENT-----PSTVSTI	1026
Db	913	LEDESLKELTKLAEVVKLDEESEDSTYBODYADPEGDNBGSNENIVKTKDITGLIV	972
QY	1027	SRNNIR-----ENVEFEASSNNI-----EVG-----SSTNEV	1054
Db	973	EPENKVKVYKHEEFTLFEANVSSSVYQNKMDHVTIVNOEAQANYEAGERKYIIONDTDE	10322
QY	1055	GSSINIESSDENIOAELGRNRGPKLMLRLGVLPREYVQSLPGSNCKPEIKKOYE	11144
Db	1033	EAHISIIERIDEN--AIGNN-----MEIPRCVCKTHNEVLE	10659
QY	1115	EVVOIV-NTDESPYLIIDNLEQPMGSSHASOVCEETPDLLDDGEIKEDTSAFEN-DIK	11717
Db	1070	RRATIENTK----ALENTNM---HDVQACQSDRDQOSTAKENEGSAKHNLDIRV	11222
QY	1172	ESSAVFSKSVOKGELSRSPPTTHLALOGVIRGAKKLESEENL-----SSDE	1221
Db	1123	SSSELESVEPLKPEBSRPNFSPPIRVGAAVVGKGVDAESFPVKKIDVMSSEDNDV	1182
QY	1222	ELPCFOHLFGKVNNIPOSQTRSHVATECSLKNTEENLS--LKNSLNDCSN--QVI--	12757
Db	1183	DIGDNDQIFKNSNTDASVNMKSVSSKERSDDEEAVIILGVYAEAINDNQNSRYINI	12424
QY	1276	-----LAKASQEHLSSEFKCSALFSFSSCSELEDITJANTNTODPFLI	1318
Db	1243	DPITNGAYEEDESEVFRQOVKKKENLHKSEEPLE-GLQSEQHFKEKDHSENEEDDTYIG	13010
QY	1319	GSSKQMRHQSESQGV-----GUSDKEFLVS---DDEERGTLGEENQOBSMDSNL--	13655
Db	1302	DITSANIHSNAPDDIKROQLLKNLSDLLENYSORLIEDSRRRKNOESDEVNTSRERDLTF	1361
QY	1366	---GEAAGCESETSVSDSCGLSSQS-----DILTQOORDMQHNLIKLOEMAEI	14144
Db	1362	EKSVNEKYAAGAIIEEDTFSE--LDISIQPREHEBDLDSNNGERSIEELNSPEAEIYEL	1419
QY	1415	EAVLE-----QHGSQPSNPS-ITDSSALED---LRNPQSTSEKAV	1454
Db	1420	E--IEGPETAASSKMNDEKORGNIPSTDLPSPDSKOEYVTSYPSYNSENMTAEKSA	14777
QY	1455	LTQKSSSEYPISONEGLSADKFEVSAASSTSKNEPEV-----ERSSPKSCPSLDDHWY	1509
Db	1478	PTSPREYEL-FSPDPNEVPMELINDEIPATLLEKHKTWTVSYLDRSHLSHDVDNEPH	1536
QY	1510	MHSCSGSLQNRRNYSQOELIYVDVEEQQLDESGPHDLETFSYLRPRODLESTPYLESGIS	1569
Db	1537	DNSINIKKNEGEEPEHQAVIDIPKVEYVEDEEMP---SKSVLEEQ-----KPSME	1584
QY	1570	LFSPD--DESPSPSED-----RAPEARVGNIPPSSTALKVYQVLKVAESAQSPAANT	1619
Db	1585	LINDKSSPENNDDETNEKDKTKAKKSKRRKRNINSRRKKKITEGSSAASNTKRRRQHE	1644
QY	1620	TDTAGV 1626	
Db	1645	PKSRGON 1651	
RESULT 12			
ANK3_HUMAN			
ID	ANK3_HUMAN	STANDARD:	PRT: 4377 AA.
AC	Q12955;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Ankyrin 3 (ANK-3) (Ankyrin G).		
GN	ANK3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain stem;		

DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104	Q105	Q106	Q107	Q108	Q109	Q110	Q111	Q112	Q113	Q114	Q115	Q116	Q117	Q118	Q119	Q120	Q121	Q122	Q123	Q124	Q125	Q126	Q127	Q128	Q129	Q130	Q131	Q132	Q133	Q134	Q135	Q136	Q137	Q138	Q139	Q140	Q141	Q142	Q143	Q144	Q145	Q146	Q147	Q148	Q149	Q150	Q151	Q152	Q153	Q154	Q155	Q156	Q157	Q158	Q159	Q160	Q161	Q162	Q163	Q164	Q165	Q166	Q167	Q168	Q169	Q170	Q171	Q172	Q173	Q174	Q175	Q176	Q177	Q178	Q179	Q180	Q181	Q182	Q183	Q184	Q185	Q186	Q187	Q188	Q189	Q190	Q191	Q192	Q193	Q194	Q195	Q196	Q197	Q198	Q199	Q200	Q201	Q202	Q203	Q204	Q205	Q206	Q207	Q208	Q209	Q210	Q211	Q212	Q213	Q214	Q215	Q216	Q217	Q218	Q219	Q220	Q221	Q222	Q223	Q224	Q225	Q226	Q227	Q228	Q229	Q230	Q231	Q232	Q233	Q234	Q235	Q236	Q237	Q238	Q239	Q240	Q241	Q242	Q243	Q244	Q245	Q246	Q247	Q248	Q249	Q250	Q251	Q252	Q253	Q254	Q255	Q256	Q257	Q258	Q259	Q260	Q261	Q262	Q263	Q264	Q265	Q266	Q267	Q268	Q269	Q270	Q271	Q272	Q273	Q274	Q275	Q276	Q277	Q278	Q279	Q280	Q281	Q282	Q283	Q284	Q285	Q286	Q287	Q288	Q289	Q290	Q291	Q292	Q293	Q294	Q295	Q296	Q297	Q298	Q299	Q300	Q301	Q302	Q303	Q304	Q305	Q306	Q307	Q308	Q309	Q310	Q311	Q312	Q313	Q314	Q315	Q316	Q317	Q318	Q319	Q320	Q321	Q322	Q323	Q324	Q325	Q326	Q327	Q328	Q329	Q330	Q331	Q332	Q333	Q334	Q335	Q336	Q337	Q338	Q339	Q340	Q341	Q342	Q343	Q344	Q345	Q346	Q347	Q348	Q349	Q350	Q351	Q352	Q353	Q354	Q355	Q356	Q357	Q358	Q359	Q360	Q361	Q362	Q363	Q364	Q365	Q366	Q367	Q368	Q369	Q370	Q371	Q372	Q373	Q374	Q375	Q376	Q377	Q378	Q379	Q380	Q381	Q382	Q383	Q384	Q385	Q386	Q387	Q388	Q389	Q390	Q391	Q392	Q393	Q394	Q395	Q396	Q397	Q398	Q399	Q400	Q401	Q402	Q403	Q404	Q405	Q406	Q407	Q408	Q409	Q410	Q411	Q412	Q413	Q414	Q415	Q416	Q417	Q418	Q419	Q420	Q421	Q422	Q423	Q424	Q425	Q426	Q427	Q428	Q429	Q430	Q431	Q432	Q433	Q434	Q435	Q436	Q437	Q438	Q439	Q440	Q441	Q442	Q443	Q444	Q445	Q446	Q447	Q448	Q449	Q450	Q451	Q452	Q453	Q454	Q455	Q456	Q457	Q458	Q459	Q460	Q461	Q462	Q463	Q464	Q465
----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

QY 61 -----CPLKNDITKRSLOESTRFSOLVEELIKIICAFOLDGTLEYANSYFAK 109  
 Db 2290 LAGLFEHKSAAVSPVHKSAAETSAQHAEKDNQMKPKLEIRI-EVHIENGQAEPTLEVIIR 2348  
 QY 110 KENNSPE-----HLKD-----EVSIIOGMYRNRAKRLLOSEPENPSLOETLSVOLSLUG 160  
 Db 2349 ETKHNPKEKMYVYOKDLISRGDINIKDLFPEKHDA---FPCSEGOOEEELTAEESLPS 2405  
 QY 161 TVRTLRTKRIORIPORTS--VYIELGSDSEEDTVANKATYCSV--GDOELLQTTPOGTREI 216  
 Db 2406 YLESSRWNTVPYQOEDSRSSQAOLISDDSYKTLKLLSHSTIEYHDELDELSELGSRARAE 2465  
 QY 217 SLSAKRAACEFSESTDYVTEHOPSNNDLNTTEKRAER--HPEK-----YOGSS-- 265  
 Db 2466 KMLSEKLDVSHSTESVTFDHAGPSPSELGSDRSREKATAPKKEILSIKYKDVSEN 2525  
 QY 266 -----VSNIHVE--PCGTNTHASSLOHENSLSL-----LTKDMN--VEKAE 303  
 Db 2526 GVGKVSXDEHFDKVTYLHSGNVSSPKHAMMRTEDRLDGRKELIYEDRVDTYKAE 2585  
 QY 304 -----FCNKSROPG---LARSQHNRMAGSKETCNDRTPTTEKKVDLNAPLCEBK 351  
 Db 2586 EKLTEVSGFFRDKTEKLNDELQSPBKARPKNGKEYSSQPTSSPEKV-LTTELASND 2644  
 QY 352 EWNK--OKLPSEN--PRDTEDPWITLNNSTIOKVEMFERSDELGSDSDHSGSESNAK 408  
 Db 2645 EWMKARQHPDQGFPPAKEEKAP--SLPSSPEK-----WVLSQOTEDSKSYTEAK 2692  
 QY 409 VADVLVDLNEVDEKSGSSEKIDLLASDPHEALICKSEVHKSVEYNEDK----- 459  
 Db 2693 GS-----ISQKADBDGPOSGFOLQKSLSTRLEFEGTAKSKDMQSEDEKSDGOSRIP 2747  
 QY 460 -----IFGTYRKKA-SLPYLS-----HYTENIIGAFYTER 490  
 Db 2748 VKKTOESKLPYQYFARKKQOKAIDLDEDESYSVOKDEMYLTKQEHQASNEI----- 2799  
 QY 491 QIIQERPLTYLKKRRRTSGLHPEDEFLK--ADLAOVKTPREMIQOG--TQOTONQGVW 546  
 Db 2800 -VVNDGSDGVNKKQRTKMSKAMPDSFSEQOKDLACHITSDLATRGWMDKVVRTWESS 2858  
 QY 547 NITSGHENTKKGSIONEKPNPI-----ESLEKESAFKTAEPDISSISNMEL-ELNT 600  
 Db 2859 GATNNKSOKEKLSHVLDVHRENIHIGPESKSVDOKEFMVSTERERKLLNGSLSEIKE 2918  
 QY 601 HNSKAPKKNRLRRKSTFRHIALFLVYVRNLSPNCTELOJSSSSSEIKKKKYNOMPV 660  
 Db 2919 MTVKSPSKKYLXR-----EYVYKEGDHPGGLD-----OPSRSESSSAVSHIPV 2962  
 QY 661 RHSRNLQMEGKEPATGAKKSNNPNEQTSKRHSDTEPE-----LKLTHNAGSFPTK 712  
 Db 2963 RVADERRMLSSNIDGFCQOAFPKHLSOKLSQSSMSKETEVTQHNSIDEKAVYSEI 3022  
 QY 713 SNTSELKEFVNPRLPREKE-----EKLEYKVVSNNAEDPKDMLSGERV-----LQ 759  
 Db 3023 SKVSKHOSYGLCPLEETETSPTRKSPDSLSEFSGKSPSSDVDPHSPIGOLEKLAPLAQ 3082  
 QY 760 TERSVESSSISLVGTQOESISLLEVSTLGAKTERP--NKCYSO-CAAFENPGLI 815  
 Db 3083 TEGSEKELIKLPIVY-----VSFVOVGQYKEKEIIOGQGVKKIISOECTVQETGTF 3132  
 QY 816 H-----GSKDNRNDETEGF-----KYPLGHEVHNSRETSIEMESELEDAQYLO 858  
 Db 3133 YTTROQOPSPQSPDPTLEQVSLDSSCKSLPTEPSSSEVSYELFNSKPTDLSL-- 3190  
 QY 859 NTFEVKSRQSALEFSNNGNAEBCATSAHSGSLKKOSPKVTECEQKLENQGNESNIK 918  
 Db 3191 -AYIPGSPRIPEVSESEEEQO-----AKSTSLKQTTVEET-AVEREMPDVSKDKNOR 3243  
 QY 919 PVQITVNTIAGPPVVGOKDPVDNAKSIKIGSRCLCSOPRGN-----ETGLTIPN 969  
 Db 3244 PKNNRVAYIERP-----PPPLD-----ADQIESDKKHNYLPKEKVDNIEVN 3285  
 QY 970 -----KHGL-----LQNPYRIPP-----LFPKISFV----- 990

Db 3286 LODEHDKQOLAEPIVIRQPPSPVPBGADVSDSDSESTIYQVPYKKYFKLEVEYDEQKE 3345  
 QY 991 --KTKCKKNLLEENFEHSHSPEREMGNENIPSYVSTISRNANNIENYKREKSSNINVG 1048  
 Db 3346 KPKASAEKASQKELIESNGSKDNEFG-LGSDSPQNEIAQNGNNDQSTITECSIATFAEFS 3404  
 QY 1049 SSTNEVG--SSINEIGSSDEN-----IOAEIGRNG-----PKL 1080  
 Db 3405 HDTATLEIDSLDGYDLODEDDGLTESDSKLEIQAMEIKKDIWNTEGLKLPADRSQSKL 3464  
 QY 1081 NAMLRGLVLOPEVYKQSLPGSNC-----KHPEIK----- 1109  
 Db 3465 EVIEEKGVGDEDEKPPSKSSSEKTPDKTDQSGAQOFTLEGHRPDRSVPPDYFSYKV 3524  
 QY 1110 KOEVEVQVQNT---DESPYLISDNL-----EQPGSSHAQVCESETPDD 1152  
 Db 3525 DEEFATPRTVATGLDFPW--SNNRQDEVPFDSKREDETKPGLAVEDRSPATYP-- 3580  
 QY 1153 LLDGELIEDTSFANDEIKESSAVFSKVGKELSRSPPTHTHLAGYRRGAKLESS 1212  
 Db 3581 -----DTPPARTIDESTPTSEP-----NFPPECKMEFMTSGA--IDMS 3620  
 QY 1213 EENLSEDEELPCFO--HLLFGK-----VNNIPSOSTRHSTVATECLSKNTEE 1238  
 Db 3621 KRDP--VEERLOFPOIGHEHTEGKSGDQEGDKSWATAPQPOSGDPTVETN-LEARNET 3677  
 QY 1259 NLSLSKSL---NDCSNQVILAKASOEHLHSEFTKCSLSFSSQCSLELDTANTNODP 1315  
 Db 3678 PVEPNPSPISPGSC-----OEGTSSSLEKSA--TMTSKVP 3716  
 QY 1316 FL-----IG--SSKOMRHOSESOVGSLSDK-----ELVSDDEEGTGL- 1351  
 Db 3717 KLRTPIKMGISASATYTKKKEGPG-EITDKLAVMTSCOGLENETITMNSANQMGVPR 3775  
 QY 1352 -EENNOEBSMDNIGEAASGESETSVEDCSGLSSQSDILTTQOARDTMOHNLKL--- 1407  
 Db 3776 HEKHDFQKDNFNNNNNLDDSTIQTDNIMSNIV-LTESAVTCTEKD--NPKVSSG 3829  
 QY 1408 -----QOEMAELEAVLEQO-----HGSQPSNPSPIISDSSALEDL 1442  
 Db 3830 KKTGVLOHCVRDQKQVLEGOOKTELGIQOKKLPKATKSPDTPRPNHMSNTKASKM 3889  
 QY 1443 RNPOSTSEKAVLTS---QKSEYPIQONPGLSADKREYVADSTSKNKP--GVERRS 1497  
 Db 3890 KOVQOSEKTKALTTSSCYDVAKSRIPKVPKPD---NIIAVKACATQKOGQPERKAKOL 3946  
 QY 1498 PSKCP-SLIDRWYHSCGSLQONRNPQOEBELIKVVDVEQOLESGPHDL-----E 1549  
 Db 3947 PSKLPAVKVRSVCVTTTTATTTTTTTTTTTTSCVVKYKQOLKEVCHSIEYFKGISGE 4006  
 QY 1550 TSYLPRDDECTPYLEGISLFSDDPPSDPEDAPASARAGNIPSSSALKVQOLKVAE 1609  
 Db 4007 TLKLVDRLESEKEMQSEL---SDEESTSRNTLSIETSR--GGQPSVYT-----KSAR 4055  
 QY 1610 SAQSPAAHNTDTGYVNAEMESVSRKPELTASTERVNKRMSVY----- 1654  
 Db 4056 DKKTAEAPLKSSEKASSEKSSRRGTQO--SPECRDIIMATVADHILGLSWTLABELN 4113  
 QY 1655 -----SGLTPEEFMLVYFARKHNTLTLNLTTEETHYVMK---TDAEVC 1697  
 Db 4114 FSVDEINOIRVENNSLISQFMFLKKWVPRDG--KNATDALTSLVTKINRIDIYTL 4190  
 QY 1698 ERLTKYFLGIAGGVVSYFVWIOSIERKMLNHEDEYKGVNNGRHHQCPKARS-- 1755  
 Db 4171 EGPITFDGNISG-----TRSFADEN-----NVFHDVPDQMGQNETSSCNLESCEA 4213  
 QY 1756 QDRKIFRGLIEICVGPPTNMPDOLFMWVOLCGASVYK-----ELSSFTLGTVPIVVO 1811  
 Db 4214 QARRVITGL-----LDRLDSPOQCRDSTISITLYKGEAGKE-ANGSH--TEIT 4258  
 QY 1812 PDANTE 1817  
 I I I I I

DB 4259 PEAKTR 4264

RESULT 13

ATRX\_HUMAN STANDARD: PRT; 2492 AA.

AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP) (Znf-HX).

GN ATRX OR RAD54L OR XH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND RP VARIANTS ATR-X.

RX MEDLINE-97123494; PubMed-8968741;

RT Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J., RA Gibbons R.J.;

RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).

RX MEDLINE-97386582; PubMed-9244431;

RT Villard L., Lossi A.-M., Cardoso C., Proud V., Chiatoni P., RA Colleaux L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE OF 860-2492 FROM N.A.

RX MEDLINE-95179111; PubMed-7874112;

RT Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V., RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E., RA Bianchi M.E., Gonzalez G.G.;

RT "Cloning and characterization of a new human Xq13 gene, encoding a putative helicase.";

RL Hum. Mol. Genet. 3:1957-1964(1994).

RN [4]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE-94214473; PubMed-8162050;

RT Gecz J., Pollard H., Gonzalez G., Villard L., Stayton C.L., RA Millasseau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";

RL Hum. Mol. Genet. 3:39-44(1994).

RN [5]

RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE-95211835; PubMed-7697714;

RT Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RL Cell 80:837-845(1995).

RN [6]

RP SEQUENCE OF 1375-2492 FROM N.A.

RA Pearce A., Chapman J.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [7]

RP MEDLINE-98167853; PubMed-9499421;

RX MEDLINE-98167853; PubMed-9499421;

RA Cardoso C., Timst S., Villard L., Khrestchatsky M., Fontes M., RA Colleaux L.;

RT "Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein.";

RL Hum. Mol. Genet. 7:679-684(1998).

RN [8]

RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC HETEROCHROMATIN.

RX MEDLINE-20040663; PubMed-10570185;

RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M., RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J., RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;

RT "Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).

RN [9]

RP DISEASE.

RX MEDLINE-20213147; PubMed-10751095;

RA Villard L., Fontes M., Ades L.C., Gecz J.;

RT "Identification of a mutation in the XNP/ATR-X gene in a family reported as Smith-Fineman-Myers syndrome.";

RL Am. J. Med. Genet. 91:83-85(2000).

RN [10]

RP VARIANT ATR-X SER-1713.

RX MEDLINE-97196774; PubMed-9043863;

RA Villard L., Lacombe D., Fontes M.;

RT "A point mutation in the XNP gene, associated with an ATR-X phenotype without alpha-thalassemia.";

RL Eur. J. Hum. Genet. 4:316-320(1996).

RN [11]

RP VARIANT JM GLN-2131.

RX MEDLINE-96224392; PubMed-8630485;

RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P., RA Munnich A., Lyonnet S.;

RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";

RL Nat. Genet. 12:359-360(1996).

RN [12]

RP VARIANTS ATR-X.

RX MEDLINE-97467722; PubMed-9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Aseunauer B., RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K., RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F., RA Higgs D.R.;

RT "Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain.";

RL Nat. Genet. 17:146-148(1997).

RN [13]

RP VARIANT ATR-X LEU-246.

RX MEDLINE-20123062; PubMed-10660327;

RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S., RA Greco D., Cardoso C., Fontes M., Ragusa A.;

RT "New mutations in XNP/ATR-X gene: a further contribution to genotype/phenotype relationship in ATR/X syndrome.";

RL Hum. Mutat. 12:214-214(1998).

RN [14]

RP VARIANT SHS LYS-1742.

RX MEDLINE-99347960; PubMed-10417298;

RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C., RA Prieto F., Fontes M., Martinez F.;

RT "Mutation of the XNP/ATR-X gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation demonstration that the mutation is involved in the inactivation bias.";

RL Am. J. Hum. Genet. 65:558-562(1999).

RN [15]

RP VARIANT CMS THR-2050.

RX MEDLINE-99326061; PubMed-10398237;

RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M., RA Curtis M.;

RT "Carpenter-Waziri syndrome results from a mutation in XNP.";

RL Am. J. Med. Genet. 85:249-251(1999).

RN [16]

RP VARIANTS ATR-X E-175; 178-V--K-198 DEL; S-190; P-219; L-246 AND C-249.

RX MEDLINE-99219535; PubMed-10204841;

RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Beloungue J., RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M., RA Lacombe D., Hanauer A., Philipp N., Schwartz C.E., Fontes M.;

RT "Evaluation of a mutation screening strategy for sporadic cases of ATR-X syndrome.";

RL J. Med. Genet. 36:183-186(1999).



```

Db 1033 KOIKNGTDDGKKSKKIRKDTSKKKDELSDYAEKSTGKGDCSDSDKSKNGAYGREKK 1092
Oy 1256 TEENLLSLKNSLNDGNOVILAKASOEHHLSEETKCSA---LFSSOCSELEDLFNTNT 1312
Db 1093 RCKLLGSKSRKRODCS-----SDPEKTSMKEDGCGNSSKRLKRIELERRLSSKRT 1146
Oy 1313 ODPFLIGSSKORHSE-----SOGVGLSDKELVSDDEERGTLSENNQESMDSNLGEA 1368
Db 1147 KRIQSGSSSSDAEESSEDNKKKQRTSSKKKAVIYEKKRNSLRSTKRQADITSSSSS 1206
Oy 1369 ASGCSEFNSVSDSCSLSSQSDILTTQQDPTMHNLIKLOQEALEAVLEQHSQP-SN 1427
Db 1207 DIEDDDQNSIGEGSS-----DEQIKIPYENLV-----LSHTGFCQSSGDALSK 1252
Oy 1428 SYPSTLSDSALDELNPEOSTEKAVLTSOKSESEPISONPGLADKFEVADSTSK 1487
Db 1253 SYPVYVDD---DDNDPENRIAKKMLLEIKAN---LSDEGSSDDEEESKRTGKO 1306
Oy 1488 NKE-PGVER-----SSPSKCPSLDRMYHNSGSLQNRNPSQELIKVVDY 1534
Db 1307 NEENPGDEFAKKNQVNSESDSESKKPRYRHLRLHK-----LVVSDG 1350
Oy 1535 EEOQLEESGPHDLTETSTYPR-----QDLEGTYLESGIS--LFSDPESDPEEDRAPES 1587
Db 1331 ESGEKKTTPKEHKEKVGKRRNRKVSSESDSEDFQSGVSEEVSESEDQRPTRSAKKA 1410
Oy 1588 ARVGNIPSTSLAKVPOLKVAESAQSPAATHTDTAGYNAMEESVREKPE 1638
Db 1411 ELEENQRSYKOKKKRRIRIKVQEDSSSEKNSNSEEEKEKEEEEEEEEEE 1461

RESULT 14
P531_HUMAN STANDARD; PRT: 1972 AA.
ID AC 012888;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tumor suppressor p53-binding protein 1 (p53-binding protein 1)
GN (53BP1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Skeletal muscle;
RX MEDLINE=94421532; PubMed=9748285;
RA Iwabuchi K., Li B., Massa H.F., Trask B.J., Date T., Fields S.;
RT "Stimulation of p53-mediated transcriptional activation by the
RT p53-binding proteins, 53BP1 and 53BP2."
RL J. Biol. Chem. 273:26061-26068(1998).
RN [2]
RP SEQUENCE OF 946-1972 FROM N.A.
RX MEDLINE=94286584; PubMed=8016121;
RA Iwabuchi K., Bartel P.L., Li B., Marracino R., Fields S.;
RT "Two cellular proteins that bind to wild-type but not mutant p53."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6098-6102(1994).
CC -1- FUNCTION: ENHANCES P53-MEDIATED TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBUNIT: BINDS TO THE CENTRAL DOMAIN OF P53.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. BOTH NUCLEAR AND CYTOPLASMIC
CC IN SOME CELLS.
CC -1- SIMILARITY: CONTAINS 2 BRC2 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: AF078776; AAC62018.1;
DR EMBL: U09477; AAA21596.1;
DR Genew: HGNC:11999; TP53BP1.
DR MIM: 605230;
DR InterPro: IPR001357; BRC2.
DR Pfam: PF00553; BRC2.
DR SMART: SM00292; BRC2.
DR PROSITE: PS50172; BRC2.
KW Nuclear protein; Transcription regulation; Activator; Repeat.
FT DOMAIN 1724 BRC2 1.
FT DOMAIN 1864 BRC2 2.
FT DOMAIN 1642 POLY-SER.
FT DOMAIN 1760 POLY-GLU.
SQ SEQUENCE 1972 AA; 213573 MW; 13E2CC8A265F9D2A CRC64;

Query Match 3.3%; Score 317.5; DB 1; Length 1972;
Best Local Similarity 17.3%; Pred. No. 5.4e-06;
Matches 389; Conservative 279; Mismatches 749; Indels 831; Gaps 88;

Oy 111 ENNSPEH--LKDEVSIIQSMGYRNRKAKRLLOSEPENP----- 145
Db 23 EDSQPEQVLEDDSGSHFSLRHLPN--LQHKENPVLDVYNSNEPOTAGEERGDGNSGF 80
Oy 146 --SLQETSLS--VQSLNIGTVFLR--TKRIQPKTSVYIELGSDSS----- 187
Db 81 NEHLKENKVAADPVDSNLDTCGISIQVLEQLPQPRITSVYLGMSVESAFAVEEKEGELE 140
Oy 188 -----EDTVAKATYCSVG-----DQELLQITPQGTREISIDS 220
Db 141 QKEKEKEEDTSCNTH--SLGAEDTASQGLGVLELSQSDQDEVENTVYEVADKEQLQSVT 199
Oy 221 AKKACAFSEETDV--TNTENHQPNSNDLNTTEKRAE---RHPEYQGSVSNLHVEFG 275
Db 200 TNSGYTRLSDVQANAIKHEQGSNEDIPAEQSSKDITVAQPSK-----DVHVV--- 249
Oy 276 THTHASSLOHENSILLTKDRMNVKAEFCNKSQPGIARSOHNMAGSKETCDNDRTPS 335
Db 250 -----KEQMPPARSEDMPSPKASVA-----A 272
Oy 336 TEKKVDLMDPLCERKENKOKLPCSENPRTQEDPWTLSNLSIQKVMWESRSDELIGS 395
Db 273 MEKKEQLSAQELME--SGLIQIKSPEVALSTOEDL-----FDQSKTVSS 316
Oy 396 DD--SHDGESESNAKVADVLVLEVDYEG-----SSKIDLLASDPH----- 437
Db 317 DGCSIPSRREGGCSLASPTATLHLQLSGQSLVQDLSLNSSDLVAPSPDARSTPFI 376
Oy 438 -EALICKSERVHSKSVESNIEDKIFGKYRK----- 468
Db 377 VPSPTQEGRODKPMDTSVLSSEEGCFPOKKLQSGEVELENNPPLLPESTVSPQASTPI 436
Oy 469 -----ASLPLNSHVTEMLITGAFVTEPQIIGR-----PLTKKLKRK 506
Db 437 SOSTVFPFPGSLPISQPOFSH-----DIFIPSLSEOSNDGKKGDMHSSSLVEEC 489
Oy 507 RPTSGLHPEDFTFKADLAV-----OKTPEMINGNOTQONQO----- 544
Db 490 SKTSELPEPNSEDGLSLTGDSCKLMLSTSEYSGPMESLSHRIDEDENQIODETE 549
Oy 545 -----VNITNSG-----HENKTKGDSIQNEKN----- 567
Db 550 PMSVPLNSKFVPAENDSLMNPADQGEVOLQSDNDKDTGDDTDRDISILATGCKGREG 609
Oy 568 -----PNPIELSEKSAFKRAEISSISNME--TELNHN----- 602
Db 610 TVAEVDCIDLTGSSSQAVPSP-----ATRSEALSSVLQOEEAMELKEHHPEGS 659
Oy 603 -----SKAFKKNLRKRSSTRIHLE-----LVVSRNLSPNCTE----- 638
Db 660 SGEVEEIPETPCESQGELEKENNESVPLHLISLITGQSGCLQCKEMPKKECEANVE 719
Oy 639 ---LOIDSCSS---SEIKKKKKYN--OMPVHRHSRLQLMGKRPATGAKSNKPNQOTS 689

```



Db 720 TVSIVSDSPQKAILDQLELHKEQEAEMEATSSEDSSVIVDVKEPSPRVVDSCEPLEGVE 779  
 QY 690 KRHDSDTFPEIKLTNAPGSFTKCSNTSELKEFVNPSLPREKEKLETVKVSNNADPK 748  
 Db 780 KCDSDSWMEDI---APEIEPCAEINLDTRKEEKSVYEGDLKSGTAEETPEVEDSSQPSL 835  
 QY 749 ----DLMLSGERVLOTERSVSSSISLVPQGDYGTQESISL-LEVSTLGKATPEBNKC 802  
 Db 836 PLVRADDPRLRLDELQOPOQTEKTSNLSLEDKSMANAKOLSSDAEAKQKGPAAHNSQSR 895  
 QY 803 SQCAA-----FENPKLIGCCSKDNNDIEGKYP-----LGH-----EYNSRETS 844  
 Db 896 CESSSTPHFTLPK-----EGDIIPLTGATPRLIOLHLEKRPKHSSTPIG 941  
 QY 845 IEMEESLDAOYLQNTFKVSKROSFALFSN-----PGNAEECATPSAHS 889  
 Db 942 I-----SNPESTIATSDVWSESMVETHDPILOSGKGDCAAPDVDDKLCRLKLVLS 993  
 QY 890 GSLKOSPVYTECECEKEENOGKN-----ESNKEPVQTVNITAGFPVVGQKD--KPVN 941  
 Db 994 PETEASEESLQFLKEKPATGERKNGSTAVAESVAPQKTMVSLVSCCEARQENEARSEDP 1053  
 QY 942 AKCSIKGSRFLCSSQFRNGTGLITPKNHGLQNPRIPLPEIKSEYKTKCKK----- 996  
 Db 1054 PTPPIRGNLHPPSSQ--GEEKEKLEGDHTTIRSOQOPKPISPVKDPVSPASQKMWIOG 1111  
 QY 997 -----NLLEENFEHSMSPER-----EMGNEN-----IPSTVSTIS 1027  
 Db 1112 PSSPOEAMVTVLEQKRGRTNKENPNSKALIERSONNIGIQMECSLRPEVTSANT 1171  
 QY 1028 RNNIRENVEKASSNINIEVSGSTNEVGSSINEIGSSDENIOALEGRNPKPLNMLRLG 1087  
 Db 1172 Q-----TIKWCEQGTSTVD-----QNFQKQATVQTERSGSKRP----- 1206  
 QY 1088 VLOPEYKOSLQSGNCKHPKIKQETEEVYQVYNDPFLYLSNLEQMGSSHASQVCS 1147  
 Db 1207 -----VSAPGDDTE--SLHSQGEET-----DMQPPHG--HYLHRHM 1240  
 QY 1148 EPPDLDLDGELKEDTSFAMNDIKESSAVESKSVQKGLSRSPSPPTHHLAGYRGAK 1207  
 Db 1241 RT-----IREVRLTVTKYI-----TDVYYVDGTEVERK 1268  
 QY 1208 KLESSENLSEDEELPCFQHLHFGKVNIPSOSTRHSTVATECLSKNTENLLSKNSL 1267  
 Db 1269 VTEETEEPI-----VECCQCEFEVPSQGTGSS 1296  
 QY 1268 NDCSN-QVILAKASQEHLSSEETKCSASFSSQCS-----ELEDLTANTNTDPELIGSS 1321  
 Db 1297 GDLGDISFSKASSLHRTSSGTSLSA-MHSSGSSGKAGAPLKGKTSSTEPADFALPSSR 1355  
 QY 1322 KQMRHOSQGVGLSKDELVSDEERGTLLENNEQSGMDSNLGAAGCSSETSVSD 1381  
 Db 1356 GGPBKISPKKVSQITPVCEDGDAGLCIRG-----GKAPVTR 1396  
 QY 1382 CGSLSSQDILTTQORDT-----MOHNL-----IKIQEAMAELEAVL 1418  
 Db 1397 GGRGRGRPSRTTGTRETAVPGLIEDISPNLSPODKSFSRVYPRVPDSTRTRIDGACA 1456  
 QY 1419 EOHGQSPMSYPSIISDSALDLRNPDESTSEKAVLITQKSEY----- 1463  
 Db 1457 LRSDSPETLFGAAGPSDGL-DASSPGNSFVGLRVAAWSSNGFYSGKITRDVAGKRY 1515  
 QY 1464 -----PISQNEGLSADFEVSADSSSTSKNEPG-----VR 1495  
 Db 1516 KLLFDGVECDVLGKDIILCDPLPDTETALSEDEY-FSAGVYKCHRESEGLYYSIEK 1574  
 QY 1496 SPSKCPSLDDRWYMH-SCSGSLQNN-----YPSQELIKVVDVEQOLEE--- 1541  
 Db 1575 ECGRK-----WYKMAVILSLQGNRLREQYGLAPYAVPPLTKAADISLNDLVEGR 1627  
 QY 1542 -----SGPHDLTETSLVRODLEGPIYLESGL-----SLFSQDPSDPSSEDRAPS 1587  
 Db 1628 KRNSNVSPATYASSSSSTTPTRKLTESPRASMGVLSGKRRLITSEERSPAK-NGRKS 1686

QY 1588 ARV-----GNIPSTSAKLPQKLVKVAESKQSPAANTDTAGY----- 1625  
 Db 1687 ATYKPAVAGAEVSPCESDNDTGEPSAL-----EEGRGLPLINKTILFLCYATLLM 1738  
 QY 1626 -NAMEESVREK-PE-LTASTERVNRKMSWVSGLTPPEEFMLVYKARKHHTLTLNLTPE 1682  
 Db 1739 ATTSDKLASKRSKLPDQPTSSSE-----EEFLTEIIPNNKQ----- 1775  
 QY 1683 ETHVAVMKDAEVC-----RLKPLFLGAGKRWVSYRWYNO 1721  
 Db 1776 --TESQLRAGAGYILDFEQAQCNFAYQCLLADQHCRTKRFKFLCLASGIPCVSHVW-- 1831  
 QY 1722 SIKERKMLNHEDEVGADVNGRHHQCP-----KRARSODRKILFRGLEICGYPPTN 1774  
 Db 1832 -----HDSCHANQLONTRNLYLLPAGYSLSEORILLMDQPRE-----NPFON 1871  
 QY 1775 MP-----TDQLEWNVOLC-----GASVYKELSSFT-----LGTGVRPIYVQPDWATEDN 1819  
 Db 1872 LKVLVSDQOQNFLELMSILMTGAAVSKQHHSSAHNKMDIALGVDPVYVTPDSC--PA 1928  
 QY 1820 GPHATQKCEAPVYTRRWVLDVVALYQC 1847  
 Db 1929 SVLKCAEALQPLPVSOEWI-----QC 1950  
 Db 1929 SVLKCAEALQPLPVSOEWI-----QC 1950  
 RESULT 15  
 ALM1\_SCHPO  
 ID ALM1\_SCHPO STANDARD; PRT; 1727 AA.  
 AC Q9UTK5; O13313; Q9UTR8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Abnormal long morphology protein 1 (Sp8).  
 GN ALM1 OR SPAC1486.04C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; Pubmed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Wolchert G., Aert R., Robben J., Crympey B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleux V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RA "The genome sequence of Schizosaccharomyces pombe."  
 RT Nature 415:871-880(2002).  
 RN [2]  
 RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=972;

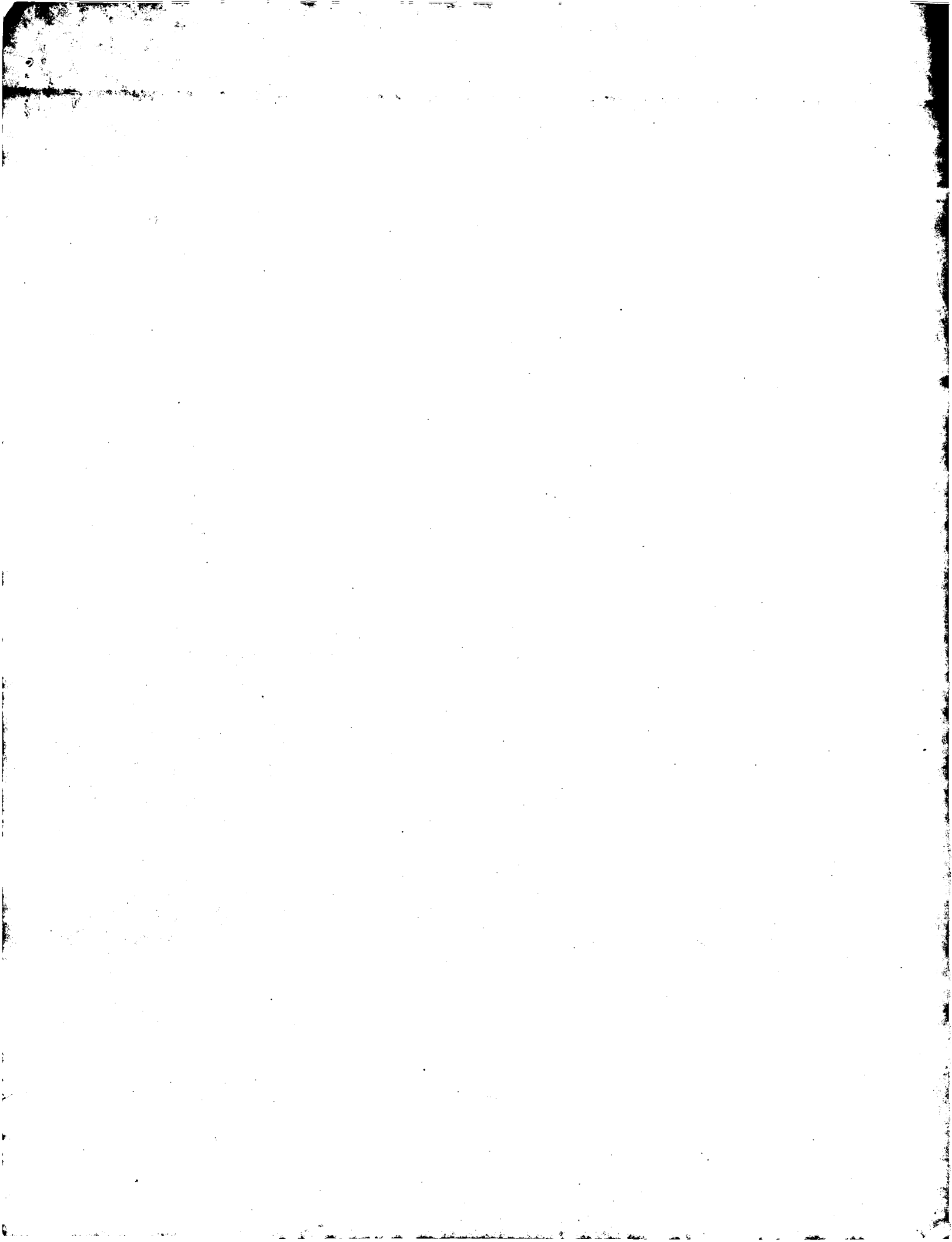


```

Db      1426 NKPSAIPATTOSEPSTVSLLEFN-STKEELSTÖRKLEIND----- 1467
QY      1290 KCSASLFSSQCELEDLTANTNTODPELIGSSK-----QMRHOSQGVGLSDKEIV 1341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1468 ----ILNTKELEKVRQNSNKS-----GTSKDTFLPNEEMERKKVMQOEVLRLRSRI 1518
QY      1342 SDEBERGTGLENNQEBQSDSNLGEASGCESETSVS----- 1379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1519 AKELQKNEILLRKONQVLODOVKALQETVVSSEFEASASVHADTKDLENLKTTEMLSVTF 1578
QY      1380 ----EDCGSLSSQSDILFT-----QQRDTMQHN----- 1403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1579 QVIFNESISDFSTSTADFTTFVOKEMEKREILQKDVEDQVQASHQKOLDNIRKLEMRN 1638
QY      1404 ---LIRLOQEMAELEAVLEQHGQSOPSNYSYPSIISDSSALEDLRNPPOSTSEKAVLTSO-- 1458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1639 KKLKSLMEKLEKLAIVRALEBQSKKDS---PAIIS---LEASKNTDSNKSNSSEVPAAQVK 1691
QY      1459 ----KSSEYPISONPEGLSADK-FEVSADSSTSK 1487
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1692 EKKLIAKTHSVDTNPPKRRSSPDAGMDVSNVKKAK 1727

```

Search completed: June 27, 2003, 10:23:22  
 Job time : 52 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: June 27, 2003, 10:18:51 ; Search time 123 Seconds  
(without alignments)  
3120.866 Million cell updates/sec

Title: US-09-734-672-4  
Perfect score: 9649  
Sequence: 1 MSLALRVEVNVNINAMOK.....LYQCELDPTYLIPQIPHSY 1863

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	9471	98.2	1863	6	09GKK8
2	9233.5	95.7	1792	4	015129
3	5795	60.1	1141	6	046485
4	5795	59.9	1141	6	046484
5	5671	58.8	1141	6	046486
6	5443.5	56.4	1140	6	046487
7	5014.5	52.0	1146	6	046488
8	4885	50.6	1812	11	0921D2
9	4873.5	50.5	1817	11	054952
10	4386.5	45.5	1146	6	046490
11	4040	41.9	1141	6	046489
12	3693	38.3	1063	11	0924E0
13	3557.5	36.9	947	6	08WMT3
14	3555.5	36.8	947	6	09BDV3
15	3482	36.1	942	6	09BDV4
16	3479	36.1	973	11	0924E1

17	3452	35.8	944	6	09BDV2	09bdv2 lama glama.
18	3435	35.6	959	11	0924E3	0924e3 sciurus nig
19	3419.5	35.4	963	11	0924MX	0924mx glaucurus v
20	3406.5	35.3	931	6	09BDW2	09bdw2 chaetophrac
21	3382	35.1	948	6	09BDU6	09bdu6 fellis silve
22	3361	34.8	942	6	09BDU8	09bdu8 physeter ca
23	3356	34.8	940	6	09BDU7	09bdu7 megaptera n
24	3345	34.7	759	4	092897	092897 homo sapien
25	3337.5	34.6	931	6	09BDU4	09bdu4 trichechus
26	3336.5	34.6	949	6	09T62	09t62 hipposidero
27	3318	34.4	934	6	08SOG2	08sg2 natalus str
28	3298	34.2	946	6	09BDV0	09bdv0 sus scrofa
29	3294	34.1	938	6	09BDW0	09bdw0 bradyptus tr
30	3293	34.1	924	6	09BDU5	09bdu5 dugong dugo
31	3283.5	34.0	945	6	09BDW1	09bdw1 tamandua te
32	3276	34.0	936	6	09T68	09t68 myotis daub
33	3263	33.8	942	6	09BDV5	09bdv5 manis sp. o
34	3262.5	33.8	931	6	08WMT8	08wmt8 solalia flu
35	3245	33.6	930	6	09BDU3	09bdu3 loxodonta a
36	3236	33.5	906	6	08SOF7	08sot7 rhinopoma h
37	3234	33.5	928	6	09BDU2	09bdu2 elephas max
38	3226	33.4	928	6	09BDU9	09bdu9 hippopotamu
39	3219	33.4	944	6	09BDV7	09bdv7 scallopus aq
40	3217	33.3	930	6	09T67	09t67 tadaria br
41	3197	33.1	928	6	09BDW3	09bdw3 dasyptus nov
42	3171.5	32.9	917	6	08WMT4	08wmt4 rousettus ra
43	3171	32.9	935	6	09T63	09t63 rhinolophus
44	3167	32.8	930	6	08SOG1	08sg1 oryctolopus
45	3157.5	32.7	929	6	09BDT4	09bdt4

## ALIGNMENTS

RESULT 1	09GKK8	PRELIMINARY	PRT; 1863 AA
ID	09GKK8		
AC	09GKK8		
DT	01-MAR-2001 (TREMUREL. 16, Created)		
DT	01-MAR-2001 (TREMUREL. 16, Last sequence update)		
DT	01-JUN-2002 (TREMUREL. 21, Last annotation update)		
DE	BRCA1 (Fragment).		
GN	BRCA1.		
OS	Pan troglodytes (Chimpanzee).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.		
OX	NCBI_TaxID=9598;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Blood:		
RA	Takeda R., Hink R.L., Jogodka C., Walter N.A.R., Messier W.;		
RT	Positive selection on the human BRCA1 gene may have resulted from		
RT	Pressure for prolonged care for infants."		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.		
DR	EMBL: AF207822; AAC43492.1;		
DR	InterPro: IPR001357; BRCT.		
DR	InterPro: IPR002378; BRCT_cancer1.		
DR	InterPro: IPR001990; Granin.		
DR	InterPro: IPR001841; Znf_fing.		
DR	Pfam: PF00533; BRCT; 2.		
DR	Pfam: PF00097; zf-C3HC4; 1.		
DR	PRINTS: PR00493; BRSTCANCER1.		
DR	SMART: SM00292; BRCT; 2.		
DR	SMART: SM00184; RING; 1.		
DR	PROSITE: PS50172; BRCT; 2.		
DR	PROSITE: PS00422; GRANINS_1; UNKNOWN_1.		
DR	PROSITE: PS00518; ZF_RING_1; 1.		
KW	Zinc-finger.		
FT	309		K -> E.
FT	590		S -> G.
FT	731		E -> K.
FT	1100		E -> G.
FT	VARIANT		

FT NON\_TER 1863 1863  
SQ SEQUENCE 1863 AA; 207916 MW: A365EB5A34A77F4A CRC64;  
Query Match 98.2%; Score 9471; DB 6; Length 1863;  
Best Local Similarity 98.3%; Pred. No. 0;  
Matches 1832; Conservative 9; Mismatches 22; Indels 0; Gaps 0;

QY 1 MDLSALRYEYOVNTINAMOKILECPICLELLEKEPVSTKCDHIFCKEPMKLINQKKGPSQ 60  
DB 1 MDLSALRYEYOVNTINAMOKILECPICLELLEKEPVSTKCDHIFCKEPMKLINQKKGPSQ 60  
QY 61 CPLCNDITKRSLOSTRFSQLYEBELKICAFOLDTGLEYANSTYFAKKNENSPHLD 120  
DB 61 CPLCNDITKRSLOSTRFSQLYEBELKICAFOLDTGLEYANSTYFAKKNENSPHLD 120  
QY 121 EYSIIOSMGYRRARARLLOSEPENPSLOETSLSVOLSNLGTVRTLTGORIOPKQSVYI 180  
DB 121 EYSIIOSMGYRRARARLLOSEPENPSLOETSLSVOLSNLGTVRTLTGORIOPKQSVYI 180  
QY 181 ELGSDSBDYVKATYCSVGDELLQITPQGTREISLDSAKKAACEFSETDVTNETHQ 240  
DB 181 ELGSDSBDYVKATYCSVGDELLQITPQGTREISLDSAKKAACEFSETDVTNETHQ 240  
QY 241 PSNNDLNTTEKRAARHPEKYOGSSVSNLHVEPCGNTYASSLOHENSLLTKDMNVE 300  
DB 241 PSNNDLNTTEKRAARHPEKYOGSSVSNLHVEPCGNTYASSLOHENSLLTKDMNVE 300  
QY 301 KAEFCNKSKOPGLARSOHNRWAGSKETCDRRTPTSEKRVDLNADPLCRKEKNKOKLPC 360  
DB 301 KAEFCNKSKOPGLARSOHNRWAGSKETCDRRTPTSEKRVDLNADPLCRKEKNKOKLPC 360  
QY 361 SENPRDTEVPWITLNSIOKVNEWFSRSDDELLGSDSHDGESESNKAVADYLDVINEVD 420  
DB 361 SENPRDTEVPWITLNSIOKVNEWFSRSDDELLGSDSHDGESESNKAVADYLDVINEVD 420  
QY 421 EYSGSSEKIDLLASPHHEALICKSERVHSKSVESNEDKIFGKYRKRAKSLPLSHVTN 480  
DB 421 EYSGSSEKIDLLASPHHEALICKSERVHSKSVESNEDKIFGKYRKRAKSLPLSHVTN 480  
QY 481 LIIGAFVTEPOITIOBRPLTNKLKRRKRPSTGLHPDEFIKKADLAOVKTEMINOGNOTE 540  
DB 481 LIIGAFVTEPOITIOBRPLTNKLKRRKRPSTGLHPDEFIKKADLAOVKTEMINOGNOTE 540  
QY 541 QNGQVMTNTSGHENKTGDSIONENKPNPIESLEESAFTKAEPDISSISNMELENI 600  
DB 541 QNGQVMTNTSGHENKTGDSIONENKPNPIESLEESAFTKAEPDISSISNMELENI 600  
QY 601 HNSKAPKKNRLRKSTRIHALVELVSNLSPPNCTELOIDSCSSSEIKKKKNOMY 660  
DB 601 HNSKAPKKNRLRKSTRIHALVELVSNLSPPNCTELOIDSCSSSEIKKKKNOMY 660  
QY 661 RHSRNLQIMEGKEPATGAKSKNKPNEOTSKRHSDTPPELKLTLNAPGFTKCSNTELKE 720  
DB 661 RHSRNLQIMEGKEPATGAKSKNKPNEOTSKRHSDTPPELKLTLNAPGFTKCSNTELKE 720  
QY 721 FVNPSLPREEBEKLETVKVSNNADPKDMLSGERVLOTERTSVSSSISLVPGTIDYGO 780  
DB 721 FVNPSLPREEBEKLETVKVSNNADPKDMLSGERVLOTERTSVSSSISLVPGTIDYGO 780  
QY 781 EYSISLEVSTIGKATERNKCYSCAFAENPKGLIHGSKDNNDTEGKYYLGHVNH 840  
DB 781 EYSISLEVSTIGKATERNKCYSCAFAENPKGLIHGSKDNNDTEGKYYLGHVNH 840  
QY 841 RETSIEMESELDAOYLONTFVSKROSFALFNSNGNMEECATSAHSGSLKKOSPKYT 900  
DB 841 RETSIEMESELDAOYLONTFVSKROSFALFNSNGNMEECATSAHSGSLKKOSPKYT 900  
QY 901 FECDKEENQCKNESNIKPVQVTNITAGFPVVGQDKPDVNAKCSIKGSRFCLSSQFNG 960  
DB 901 FECDKEENQCKNESNIKPVQVTNITAGFPVVGQDKPDVNAKCSIKGSRFCLSSQFNG 960  
QY 961 NETGLITPNKGLLONPYRIPPLPIKSFVTKCKKNLLENFEEHSMSPEREMGENIP 1020  
DB 961 NETGLITPNKGLLONPYRIPPLPIKSFVTKCKKNLLENFEEHSMSPEREMGENIP 1020

DB 961 NETGLITPNKGLLONPYRIPPLPIKSFVTKCKKNLLENFEEHSMSPEREMGENIP 1020  
QY 1021 STVSTISRNINIRENVFKKASSNINEVGSSTNEVGSSINIEGSSPDENIOAELGRNRGPKL 1080  
DB 1021 STVSTISRNINIRENVFKKASSNINEVGSSTNEVGSSINIEGSSPDENIOAELGRNRGPKL 1080  
QY 1081 NAMRLGLVLOPEVYKOSLPGSNCKHPEIKKOEEVQVTNMTDPSYLLISDNLEQPMGSS 1140  
DB 1081 NAMRLGLVLOPEVYKOSLPGSNCKHPEIKKOEEVQVTNMTDPSYLLISDNLEQPMGSS 1140  
QY 1141 HASQVSETPDLDLDDGETKEDTSAENDIKESSAVFSKSVQKGLSPPFTHTHLAQ 1200  
DB 1141 HASQVSETPDLDLDDGETKEDTSAENDIKESSAVFSKSVQKGLSPPFTHTHLAQ 1200  
QY 1201 GYRGAKKLESSEENLSDEELPCFOHLIFKAVNNISQSOGRHSTVATECLSKNTEENL 1260  
DB 1201 GYRGAKKLESSEENLSDEELPCFOHLIFKAVNNISQSOGRHSTVATECLSKNTEENL 1260  
QY 1261 LSLKNSLNDCSNOVILLAKASQEHLSSEETKCSASLFSQCSLEDELDTANTMTQDPFLIGS 1320  
DB 1261 LSLKNSLNDCSNOVILLAKASQEHLSSEETKCSASLFSQCSLEDELDTANTMTQDPFLIGS 1320  
QY 1321 SKQMRHQSOGVGLSDKELVSDDEERGTLSENNQESQSDMSNLGEAASCESETSYSE 1380  
DB 1321 SKQMRHQSOGVGLSDKELVSDDEERGTLSENNQESQSDMSNLGEAASCESETSYSE 1380  
QY 1381 DCSGLSSQSDILTTOOROTMOHNLJKLOEAMELFAVLEOHQSOPSNSTPSTIDSQSALE 1440  
DB 1381 DCSGLSSQSDILTTOOROTMOHNLJKLOEAMELFAVLEOHQSOPSNSTPSTIDSQSALE 1440  
QY 1441 DLNRPQOSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSAADSSTSKNKEGVERSSPSK 1500  
DB 1441 DLNRPQOSTSEKAVLTSOKSSEYPISONPEGLSADKFEVSAADSSTSKNKEGVERSSPSK 1500  
QY 1501 CPELDRWYMHSCSGLQNTNPSQEEILKYVDVEEQLESGBPDLTETSYLRQDLEG 1560  
DB 1501 CPELDRWYMHSCSGLQNTNPSQEEILKYVDVEEQLESGBPDLTETSYLRQDLEG 1560  
QY 1561 TPYLESGISLFDDESDPSEDEKAPESAHVGNIPSTSAKVPOLKVAESQSPAHAHTT 1620  
DB 1561 TPYLESGISLFDDESDPSEDEKAPESAHVGNIPSTSAKVPOLKVAESQSPAHAHTT 1620  
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSNVSGLTPPEEMLYYKFAKRNHITLNL 1680  
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSNVSGLTPPEEMLYYKFAKRNHITLNL 1680  
QY 1681 TREETTHVVKTTDAEFVCERTLKYFLGIGAKWVSYFWVTOSIKERKMLNHDREVGDV 1740  
DB 1681 TREETTHVVKTTDAEFVCERTLKYFLGIGAKWVSYFWVTOSIKERKMLNHDREVGDV 1740  
QY 1741 VNGRNHOGPKRARESDRKIFRGLICCYGPTNNPTQOLBMMVQOLGASVYKELSSFTL 1800  
DB 1741 VNGRNHOGPKRARESDRKIFRGLICCYGPTNNPTQOLBMMVQOLGASVYKELSSFTL 1800  
QY 1801 GTGVHPVIVVQPDANTEDNGFHAIGQMCAPVYTRENVLDSVALYQCOELDTYLIPOIPH 1860  
DB 1801 GTGVHPVIVVQPDANTEDNGFHAIGQMCAPVYTRENVLDSVALYQCOELDTYLIPOIPH 1860  
QY 1861 SHY 1863  
DB 1861 SHY 1863

RESULT 2  
ID 015129 PRELIMINARY; PRT: 1792 AA.  
AC 015129;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)  
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)  
GN BRCA1.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RX MEDLINE=96172829; PubMed=8569721;  
 RA Holt J.T., Thompson M.E., Szabo C., Robinson-Benton C., Arteaga C.L.,  
 RA King M.C., Jensen R.A.;  
 RT "Growth retardation and tumour inhibition by BRCA1.";  
 RL Nat. Genet. 12:298-302(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Holt J.T., Robinson-Benton C.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF005068; AAB61673.1; -;  
 DR InterPro: IPR001357; BRCT.  
 DR InterPro: IPR002378; Bist\_cancer1.  
 DR InterPro: IPR001990; Granin.  
 DR Pfam: PF00533; BRCT; 2.  
 DR PRINTS: PR00493; BRSTCANCER1.  
 DR SMART: SM00292; BRCT; 2.  
 DR PROSITE: PS00172; BRCT; 2.  
 DR PROSITE: PS00422; GRANINS\_1; UNKNOWN\_1.  
 SQ SEQUENCE 1792 AA; 199692 MW; 0CB9B06C0791AA44 CRC64;

Query Match 95.7%; Score 9233.5; DB 4; Length 1792;  
 Best Local Similarity 99.8%; Pred. No. 0;

Matches 1788; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

72 SLOESTRFQOLVBEELKICAFOLDGLGYANSYNFAKKENNSPEHLKDEVTIISGMGR 131  
 2 SLOESTRFQOLVBEELKICAFOLDGLGYANSYNFAKKENNSPEHLKDEVTIISGMGR 61  
 132 NNAKRLLOSEPENPSLOETSLSVOLSLGTVTLRTKRIQPKTSVYIELGSDSEDTY 191  
 62 NNAKRLLOSEPENPSL-ETSLSVOLSLGTVTLRTKRIQPKTSVYIELGSDSEDTY 120  
 192 NKATYCSVGDDELOTTPOGTDEISLDSAKKAACEFSETDVTNHHOPSNNDLTTEK 251  
 121 NKATYCSVGDDELOTTPOGTDEISLDSAKKAACEFSETDVTNHHOPSNNDLTTEK 180  
 252 RAERHPEKYQSSVSNLHVEPCGTNTNTHASSLQEHENSSLLTKDRNNVKAKEFCNKSOP 311  
 181 RAERHPEKYQSSVSNLHVEPCGTNTNTHASSLQEHENSSLLTKDRNNVKAKEFCNKSOP 240  
 312 GLARSHNHWAGSKETCNDRTPSTEKKYDLNADPLCEKRENNKOKLPCSENPROTDEYV 371  
 241 GLARSHNHWAGSKETCNDRTPSTEKKYDLNADPLCEKRENNKOKLPCSENPROTDEYV 300  
 372 WITLNSIOKVNMWFSRDELLGSDSHDGESFSNAKVADVDVLENEVDEYSSSSKIDL 431  
 301 WITLNSIOKVNMWFSRDELLGSDSHDGESFSNAKVADVDVLENEVDEYSSSSKIDL 360  
 432 LASDPHEALICKSERVHKSVESENIEDKIFGKTYRKKASLPNLSHVTENLIIIGAFTEPO 491  
 361 LASDPHEALICKSERVHKSVESENIEDKIFGKTYRKKASLPNLSHVTENLIIIGAFTEPO 420  
 492 IIOERPLTNKLKRRKRPSTGLHPEDEFIKADLAVALOKTPMINOGTNOTEQONVMTINS 551  
 421 IIOERPLTNKLKRRKRPSTGLHPEDEFIKADLAVALOKTPMINOGTNOTEQONVMTINS 480  
 552 GHEKTKGDSIONENKPNPISLEKESAFKTAEPYSSISNMNLELNTNHSKAPKKNLU 611  
 481 GHEKTKGDSIONENKPNPISLEKESAFKTAEPYSSISNMNLELNTNHSKAPKKNLU 540  
 612 RRRKSTRHIALELVYSRNLSPNCTELOIDSCSSSEELIKKKKYNOMPVHRSHNLQAMEG 671  
 541 RRRKSTRHIALELVYSRNLSPNCTELOIDSCSSSEELIKKKKYNOMPVHRSHNLQAMEG 600  
 672 KEPATGAKSKNPNEDOTSKRHSDTFPELKLINAPESTFKCSNTSELKEFVNSLPREEK 731

601 KEPATGAKSKNPNEDOTSKRHSDTFPELKLINAPESTFKCSNTSELKEFVNSLPREEK 660  
 732 EEKLETVKSNNAEDKDLMLSGERYLOTERSVESSTISLYPEOTDGTQESTISLEVSUTL 791  
 661 EEKLETVKSNNAEDKDLMLSGERYLOTERSVESSTISLYPEOTDGTQESTISLEVSUTL 720  
 792 GAKTEPNKCVSOCAFAFENPKGLIHGCSKDNNDTGFYKPYPLGHEVNHSHRETSIEMEES 851  
 721 GAKTEPNKCVSOCAFAFENPKGLIHGCSKDNNDTGFYKPYPLGHEVNHSHRETSIEMEES 780  
 852 LDAQVLONTPEKYSKROSFALFSPNGBAEEBCATFSAHSGSLKKQSPKVFTECEQKEBNOG 911  
 781 LDAQVLONTPEKYSKROSFALFSPNGBAEEBCATFSAHSGSLKKQSPKVFTECEQKEBNOG 840  
 912 KNESENKPVQVNTITGFPVVGOKDPVDNAKCSINGSGSFCLSSOFRGETLITPNKH 971  
 841 KNESENKPVQVNTITGFPVVGOKDPVDNAKCSINGSGSFCLSSOFRGETLITPNKH 900  
 972 GLLONPYRIPPLPEPIKSFYKCKKNLLENEFEHSMSPERENGNEIPSTVSTISRNNI 1031  
 901 GLLONPYRIPPLPEPIKSFYKCKKNLLENEFEHSMSPERENGNEIPSTVSTISRNNI 960  
 1032 RENWFEKASSNINEYGSSTNEVGSSINEIGSDENIQALGNRPKLNAMLRLGLVLP 1091  
 961 RENWFEKASSNINEYGSSTNEVGSSINEIGSDENIQALGNRPKLNAMLRLGLVLP 1020  
 1092 EYVKOSLPGSNCKHPPIKKEEYVQVNTDPSPLIDNLEQPMGSSHASVGCETPD 1151  
 1021 EYVKOSLPGSNCKHPPIKKEEYVQVNTDPSPLIDNLEQPMGSSHASVGCETPD 1080  
 1152 DILLDOEIKEDTFFAANDIKESSAVFSKSVOKELRSRSPFTHTLHAQYRGAKKLES 1211  
 1081 DILLDOEIKEDTFFAANDIKESSAVFSKSVOKELRSRSPFTHTLHAQYRGAKKLES 1140  
 1141 SEENLSSDEDELPFOHLLFGKVNNTIPSOSTRHSTVATECLSKNTEENLILKNSLNDCS 1200  
 1212 SEENLSSDEDELPFOHLLFGKVNNTIPSOSTRHSTVATECLSKNTEENLILKNSLNDCS 1271  
 1272 NOVYILAKASOEHLSETKCSASFSSOCSELDLANTOOPFLIGSSKOMRHOSEGO 1331  
 1201 NOVYILAKASOEHLSETKCSASFSSOCSELDLANTOOPFLIGSSKOMRHOSEGO 1260  
 1332 GYGLSDKELVSDDEEGTGLLENNQEOQMSDNLGEAASCESETSVEDCSGLSSQSDI 1391  
 1261 GYGLSDKELVSDDEEGTGLLENNQEOQMSDNLGEAASCESETSVEDCSGLSSQSDI 1320  
 1392 LTTQORDTQOHLNLIKQOEAELEAVLEQHGSPNSYSPIISDSSALEDLRNPPOSTSE 1451  
 1321 LTTQORDTQOHLNLIKQOEAELEAVLEQHGSPNSYSPIISDSSALEDLRNPPOSTSE 1380  
 1452 KAVLTOKSSEXPISQNPGLSADKFEVSADSTSNKKEGVRSSPSKPSLDDRWYMH 1511  
 1381 KAVLTOKSSEXPISQNPGLSADKFEVSADSTSNKKEGVRSSPSKPSLDDRWYMH 1440  
 1512 SCGSLQNNRNPQOEELIVVDVEEQOLESGPHDLTETSYLPRQDLEGPYLESGISLF 1571  
 1441 SCGSLQNNRNPQOEELIVVDVEEQOLESGPHDLTETSYLPRQDLEGPYLESGISLF 1500  
 1572 SDDPESDPSEDRAPESARVGNTPSSTALKVPOLKVAESAQSPAAAHHTDGTAGYNAMESS 1631  
 1501 SDDPESDPSEDRAPESARVGNTPSSTALKVPOLKVAESAQSPAAAHHTDGTAGYNAMESS 1560  
 1632 VSREKPELTASTERVKKRMVAVSGLTPEEPMLVYFAKHHHTLNLTEETHVYMKT 1691  
 1561 VSREKPELTASTERVKKRMVAVSGLTPEEPMLVYFAKHHHTLNLTEETHVYMKT 1620  
 1692 DAEFVCEKTLKFLGLAGAKWVSYFWYQSIKERKMLNEHDFEYGVADVNGNHHGPRR 1751  
 1621 DAEFVCEKTLKFLGLAGAKWVSYFWYQSIKERKMLNEHDFEYGVADVNGNHHGPRR 1680  
 1752 ARESDOKRTFRGLETICCPPTNMPDOLLENNVQLCGASVYKELSSFTLGTGVHPITVVO 1811  
 1681 ARESDOKRTFRGLETICCPPTNMPDOLLENNVQLCGASVYKELSSFTLGTGVHPITVVO 1740

QY 1812 PDANTDNGFHAIGMCEAPVVTREMLDVALYOCQELDTYLIPQIPSHY 1863  
 |||||||  
 Db 1741 PDANTDNGFHAIGMCEAPVVTREMLDVALYOCQELDTYLIPQIPSHY 1792

RESULT 3  
 ID 046485 PRELIMINARY; PRT: 1141 AA.  
 AC 046485:  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer  
 susceptibility protein) (Fragment).  
 GN BRCAL.  
 OS Gorilla gorilla gorilla (Lowland gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.  
 OX NCBI\_TaxID=9595;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98122577; PubMed=9462745;  
 RA Hacia J.G., Makowski W., Edgemon K., Erdos M.R., Robbins C.M.,  
 RA Fodor S.P.A., Brody L.C., Collins F.S.;  
 RT "Evolutionary sequence comparisons using high-density oligonucleotide  
 arrays.";  
 RL Nat. Genet. 18:155-158(1998).  
 CC -1- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE  
 IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 DR EMBL: AF019076; AAC39584.1; -  
 DR InterPro: IPR002378; Brst\_cancerl.  
 DR PRINTS: PRO0493; BRSTCANCERL.  
 KM DNA-binding; Nuclear protein; Anti-oncogene.  
 FT NON\_TER 1  
 FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 427 430 POLY-LYS.  
 FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).  
 FT NON\_TER 1141 1141  
 SQ SEQUENCE 1141 AA; 127339 MW; A09531803BB99C3D CRC64;

Query Match 60.1%; Score 5795; DB 6; Length 1141;  
 Best Local Similarity 98.1%; Pred. No. 3,8e-291;  
 Matches 1119; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 225 ACEFSEDTVTNTEHOPSNNDLNTKRAAEHRPEKYGSSVSNLHVEPCGNTTHASSIQ 284  
 |||||||  
 Db 1 ACESERDVYNTTEHOPSNNDLNTTEKRAETERHPEKYGSSVSNLHVEPCGNTTHASSIQ 60

QY 285 HENSLLLTCKDRMNVKAEFCNKSQKPGIARSOHNRNAGSKETCDNRTPTEKRVLDNA 344  
 |||||||  
 Db 61 HENSLLLTCKDRMNVKAEFCNKNKQGLARSQHRNAGSKETCDNRTPTEKRVLDNA 120

QY 345 DPLCEKEMKQKLPCESENPRDTEVPWITLNSIQKYNEMFSRDELLGSDSDHGESE 404  
 |||||||  
 Db 121 DPLCEKEMKQKLPCESENPRDTEVPWITLNSIQKYNEMFSRDELLGSDSDHGESE 180

QY 405 SNAKAVADVLTNVEDEYSGSSEKIDLASDPHEALIKKSEVHKSVESNIEDKIFGKT 464  
 |||||||  
 Db 181 SNAKAVADVLTNVEDEYSGSSEKIDLASDPHEALIKKSEVHKSVESNIEDKIFGKT 240

QY 465 YRRKASLPNLSHVTENLIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKADIA 524  
 |||||||  
 Db 241 YRRKASLPNLSHVTENLIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKADIA 300

QY 525 VQKTPENINQSTNTQNGVYMNITNSCHENKTKGDSIQNEKNPPIISLEKESAFKTKA 584  
 |||||||  
 Db 301 VQKTPENINQSTNTQNGVYMNITNSCHENKTKGDSIQNEKNPPIISLEKESAFKTKA 360

QY 585 EPISSISNMELNINHSKAPKKRNRKRKSTRHIAHLVLSNLSPPNCTEIQDSC 644  
 |||||||  
 Db 361 EPISSISNMELNINHSKAPKKRNRKRKSTRHIAHLVLSNLSPPNCTEIQDSC 420

QY 645 SSSSEIKKKKYNNQMFVRHSRLQMEGKEPATGAKSKNPNQDTSKRHSDTPPELKTN 704  
 |||||||  
 Db 421 SSSSEIKKKKYNNQMFVRHSRLQMEGKEPATGAKSKNPNQDTSKRHSDTPPELKTN 480

QY 705 APGSEFKGNTSELKEFVNPSPLPREEKEKETKYKVSNNADPPDLMLSGERVLYQTERSV 764  
 |||||||  
 Db 481 APGSEFKGNTSELKEFVNPSPLPREEKEKETKYKVSNNADPPDLMLSGERVLYQTERSV 540

QY 765 ESSSISLVPGTDYQGESISLIEVSTLTKATPEPKVSQCAAFENPGGLHGCSSKNRN 824  
 |||||||  
 Db 541 ESSSISLVPGTDYQGESISLIEVSTLTKATPEPKVSQCAAFENPGGLHGCSSKNRN 600

QY 825 DTEGFRYPLGHEVNSRETSIEMESELDAQYLONTFVKSRQSFALFNSGNAEECAT 884  
 |||||||  
 Db 601 DTEGFRYPLGHEVNSRETSIEMESELDAQYLONTFVKSRQSFALFNSGNAEECAT 660

QY 885 FSAHSGSLKKQSPKVTPECQKEENQGNESNIKPVQTVNTAGPVPVQKDKPVDNAKC 944  
 |||||||  
 Db 661 FSAHSGSLKKQSPKVTPECQKEENQGNESNIKPVQTVNTAGPVPVQKDKPVDNAKC 720

QY 945 STKGSRCTLSQPFQNETGLTPPKHGLLPNPRIPRIPKSPVTKCKNLLNEPFE 1004  
 |||||||  
 Db 721 STKGSRCTLSQPFQNETGLTPPKHGLLPNPRIPRIPKSPVTKCKNLLNEPFE 780

QY 1005 EHSMSPEREMGNENIPSTVSTISRNNIRENVFEKASSNINEVGSSTNEVGSSINEIGSS 1064  
 |||||||  
 Db 781 EHSMSPEREMGNENIPSTVSTISRNNIRENVFEKASSNINEVGSSTNEVGSSINEIGSS 840

QY 1065 DENIOAELGRNRPGLMAMLRGLVLPYVYKQSLDGSCKIPETIKQDEYVQTVNTDF 1124  
 |||||||  
 Db 841 DENIOAELGRNRPGLMAMLRGLVLPYVYKQSLDGSCKIPETIKQDEYVQTVNTDF 900

QY 1125 SPFLTSDNLEOPMGSSHASQVCSSETPDLLDDELKEDTSAENDIKESSAVFSKQVG 1184  
 |||||||  
 Db 901 SPFLTSDNLEOPMGSSHASQVCSSETPDLLDDELKEDTSAENDIKESSAVFSKQVG 960

QY 1185 ELRSRSPSEPTHTLTAQYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNIPISQSTRH 1244  
 |||||||  
 Db 961 ELRSRSPSEPTHTLTAQYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNIPISQSTRH 1020

QY 1245 STVAITECSKNTENELLSKNSLNDCSNQVTLAAQSEHHLSEETKCSASLFSQGELE 1304  
 |||||||  
 Db 1021 STVAITECSKNTENELLSKNSLNDCSNQVTLAAQSEHHLSEETKCSASLFSQGELE 1080

QY 1305 DLTANTNODPFLIGSSKOMRHQESQGVGLSDKELVSDDERGTGLENNQDEQSDSN 1364  
 |||||||  
 Db 1081 DLTANTNODPFLIGSSKOMRHQESQGVGLSDKELVSDDERGTGLENNQDEQSDSN 1140

QY 1365 L 1365  
 ||  
 Db 1141 L 1141

RESULT 4  
 ID 046484 PRELIMINARY; PRT: 1141 AA.  
 AC 046484:  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer  
 susceptibility protein) (Fragment).  
 GN BRCAL.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98122577; PubMed=9462745;  
 RA Hacia J.G., Makowski W., Edgemon K., Erdos M.R., Robbins C.M.,  
 RA Fodor S.P.A., Brody L.C., Collins F.S.;



RT "Evolutionary sequence comparisons using high-density oligonucleotide arrays".  
 RL Nat. Genet. 18:155-158(1998).  
 CC -!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 DR EMBL: AF019075; AAC39583.1; -.  
 DR InterPro: IPR002378; Brst\_cancer1.  
 DR PRINTS: PR00493; BRSTCANCER1.  
 KM DNA-binding: Nuclear protein; Anti-oncogene.  
 FT NON\_TER 1  
 FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 427 430 POLY-LYS.  
 FT DOMAIN 505 509 POLY-GLU.  
 FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).  
 FT NON\_TER 1141 1141  
 SQ SEQUENCE 1141 AA; 127399 MW; D7260B3CA6632D5 CRC64;

Query Match 59.9%; Score 5776; DB 6; Length 1141;  
 Best Local Similarity 97.9%; Pred. No. 3.6e-290;  
 Matches 117; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

QY 225 ACDEFSTDTNTNHHQPSNNDLNTTEKRAERHPEKYQSSVSNLHVEPCGNTTHASSIQ 284  
 DB 1 ACEFSETDVTNTEHHQPSNNDLNTTEKRAERHPEKYQSSVSNLHVEPCGNTTHASSIQ 60

QY 285 HENSLLTKRDMNNVKKAFCKSKQPGARSOHNWAGSKETCNDRPTSTKKYKDLNA 344  
 DB 61 HENSLLTKRDMNNVKKAFCKSKQPGARSOHNWAGSKETCNDRPTSTKKYKDLNA 120

QY 345 DPLCEKERNKOKLPCSENPRDTEDPWTLLNSIQKVMFESRSDSLGSDSDHGESE 404  
 DB 121 DPLCEKERNKOKLPCSENPRDTEDPWTLLNSIQKVMFESRSDSLGSDSDHGESE 180

QY 405 SNAKADVDVINEVDYSGSSEKIDLASDPHEALICSEKRVHKSQSVSNEDKIFGKT 464  
 DB 181 SNAKADVDVINEVDYSGSSEKIDLASDPHEALICSEKRVHKSQSVSNEDKIFGKT 240

QY 465 YRKASLPLSHVTENLIGAFVTEPQIOERPLTKLRKRPTSGLPDEFIKKADIA 524  
 DB 241 YRKASLPLSHVTENLIGAFVTEPQIOERPLTKLRKRPTSGLPDEFIKKADIA 300

QY 525 VQKTPMINQGTQNGQVNNITNSGHENKTGDSIONENPNPIESLEKESAFKTKA 584  
 DB 301 VQKTPMINQGTQNGQVNNITNSGHENKTGDSIONENPNPIESLEKESAFKTKA 360

QY 585 EPLISSISMELELNIHNSKAPKKNLRKSSRTHIALELVYSRNLSPNCTELQIDSC 644  
 DB 361 EPLISSISMELELNIHNSKAPKKNLRKSSRTHIALELVYSRNLSPNCTELQIDSC 420

QY 645 SSSEELKKKKYKQMPVRSRNLQIMGKREPATGAKSKSNKPNQTSKRHSDTFPELKLTN 704  
 DB 421 SSSEELKKKKYKQMPVRSRNLQIMGKREPATGAKSKSNKPNQTSKRHSDTFPELKLTN 480

QY 705 AGSFETKCSNTSELKEFVNPSPREEKEKLETVKVSNNAPDPKDLMSGERVLOTERSV 764  
 DB 481 AGSFETKCSNTSELKEFVNPSPREEKEKLETVKVSNNAPDPKDLMSGERVLOTERSV 540

QY 765 ESSSISLVPGTDYGOESISLLEVTSLGAKTEPNKCVSQAFAENPKGLHGCSKDNRN 824  
 DB 541 ESSSISLVPGTDYGOESISLLEVTSLGAKTEPNKCVSQAFAENPKGLHGCSKDNRN 600

QY 825 DREGFKYPLGHEVNSRSTISEMSESLDAOYLQNTFKSKOSFALFSPNNAEBCXT 884  
 DB 601 DREGFKYPLGHEVNSRSTISEMSESLDAOYLQNTFKSKOSFALFSPNNAEBCXT 660

QY 885 FSAHSGSLKQSPKVTFFCEQKEENQKNESNIKPVQVNTAGFPVVGQXKRPVDNAKC 944  
 DB 661 FSAHSGSLKQSPKVTFFCEQKEENQKNESNIKPVQVNTAGFPVVGQXKRPVDNAKC 720

QY 945 SIKGSGRFLSSQFQNGENGLITPNKHGLLQNPYRIPPLPIKSFYKTKCKNNLEENFE 1004  
 DB 721 SIKGSGRFLSSQFQNGENGLITPNKHGLLQNPYRIPPLPIKSFYKTKCKNNLEENFE 780

QY 1005 EHSMSPERENGENIPSTVSTISRNNIRENVFEKASSNNINEVGSSSTNEIGSS 1064  
 DB 781 EHSMSPERENGENIPSTVSTISRNNIRENVFEKASSNNINEVGSSSTNEIGSS 840

QY 1065 DENIQAEILGRNRPKILNMLRLGLVLOPEYVKOSLPSSNCKHPEIKQEEVYQVNTDF 1124  
 DB 841 DENIQAEILGRNRPKILNMLRLGLVLOPEYVKOSLPSSNCKHPEIKQEEVYQVNTDF 900

QY 1125 SPLIISDNLQPPGSSHASQVSETPDDLLDGEIKEDTSPFANDIKESAVFSKSVQNG 1184  
 DB 901 SPLIISDNLQPPGSSHASQVSETPDDLLDGEIKEDTSPFANDIKESAVFSKSVQNG 960

QY 1185 ELSRSPSPPTHTHLAGYRGAKKLSESEENLSEDEELPCPHLLFGKVSNPSTH 1244  
 DB 961 ELSRSPSPPTHTHLAGYRGAKKLSESEENLSEDEELPCPHLLFGKVSNPSTH 1020

QY 1245 STVATECLSKNTEENLISLNSLDCSNQVILAKASQEHHLSEETKCSALFSSQSELE 1304  
 DB 1021 STVATECLSKNTEENLISLNSLDCSNQVILAKASQEHHLSEETKCSALFSSQSELE 1080

QY 1305 DLTANTNDPPLIGSSKQMRHOSQGVLSQDELVSDDERGTGLENNDEQSGMSDN 1364  
 DB 1081 DLTANTNDPPLIGSSKQMRHOSQGVLSQDELVSDDERGTGLENNDEQSGMSDN 1140

QY 1365 L 1365  
 DB 1141 L 1141

RESULT 5  
 046486 PRELIMINARY; PRT; 1141 AA.  
 AC 046486;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer susceptibility protein) (Fragment).  
 GN BRCA1.  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.  
 OC NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98122577; PubMed=9462745;  
 RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M., Fodor S.P.A., Brody L.C., Collins F.S.;  
 RT "Evolutionary sequence comparisons using high-density oligonucleotide arrays".  
 RL Nat. Genet. 18:155-158(1998).  
 CC -!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 DR EMBL: AF019077; AAC39585.1; -.  
 DR InterPro: IPR002378; Brst\_cancer1.  
 DR PRINTS: PR00493; BRSTCANCER1.  
 KM DNA-binding: Nuclear protein; Anti-oncogene.  
 FT NON\_TER 1  
 FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 427 430 POLY-LYS.  
 FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).  
 FT NON\_TER 1141 1141  
 SQ SEQUENCE 1141 AA; 127450 MW; A4C692E5779091F1 CRC64;

Query Match 58.8%; Score 5671; DB 6; Length 1141;  
 Best Local Similarity 96.3%; Pred. No. 9.6e-285;  
 Matches 109; Conservative 11; Mismatches 31; Indels 0; Gaps 0;

QY 225 ACDEFSTDTNTEHHQPSNNDLNTTEKRAERHPEKYQSSVSNLHVEPCGNTTHASSIQ 284  
 DB 1 ACEFSETDVTNTEHHQPSNNDLNTTEKRAERHPEKYQSSVSNLHVEPCGNTTHASSIQ 60

QY	285	HENSLLLTCKDMANKEAFCKNSKQOPGLARQONHRMAASKETPCNDRCRPTSEKRYVDLNA	344
Db	61	HENSLLLTCKDMANKEAFCKNSKQOPGLARQONHRMAASKETPCNDRCRPTSEKRYVDLNA	120
QY	345	DPLCKEREMNKOLPCSENPREDTEVPWITLNNSSIOKVEMFSRDELLGSDSDSHDGESE	404
Db	121	DPLCKEREMNKOLPCSENPREDTEVPWITLNNSSIOKVEMFSRDELLGSDSDSHDGRSE	180
QY	405	SNAKVADVLADYNEVDEYSGSSEKIDLLASDPHEALICKSESVHKSQVSNIEDTIFGKT	464
Db	181	SNAKVADVLADYNEVDEYSGSSEKIDLLASDPHEALICKSESVHKSQVSNIEDTIFGKT	240
QY	465	YRKKASLPNLSHVTENLIIIGAFVTEPQIIQOERPLTNLTKRRRRPTSGLHPEDFIKKADLA	524
Db	241	YRKKASLPNLSHVTENLIIIGAFVTEPQIIQOERPLTNLTKRRRRPTSGLHPEDFIKKADLA	300
QY	525	VQKTPBEMINOCTNOQEOQVANNITNSGHENKTKGDSIONEKNPNPISLEKESAFKTKA	584
Db	301	VQKTPBEMINOQVANOQVANNITNSGHENKTGDSIONEKNPNPISLEKESAFKTKA	360
QY	585	EPISISSIMETELMIHNSKAPKRRRLRKSSSTRHHALEYVSRNLSPNCTELOIDSC	644
Db	361	EPISISSIMETELMIHNSKAPKRRRLRKSSSTRHHALEYVSRNLSPNCTELOIDSC	420
QY	645	SSSEIEIKKKYQMPVHRHSRNLQIMEGKEPATGAASKNKPNEOTSKRHDSDFPPELKLTN	704
Db	421	SSSEIEIKKKYQMPVHRHSRNLQIMEGKEPATGAASKNKPNEOTSKRHDSDFPPELKLTN	480
QY	705	APGSTKCSNTSELKEFEYNPSLPREKEFEKLTIVYVSNNAEDPKDMLSGEVDLQTERSV	764
Db	481	APGSTKCSNTSELKEFEYNPSLPREKEFEKLTIVYVSNNAEDPKDMLSGEVDLQTERSV	540
QY	765	ESSSISLVPQTDYQTOESTISLLEVSTLGAKTAPKACVSOCAAFENPKGLIHGSKARDNN	824
Db	541	ESSSISLVPQTDYQTOESTISLLEVSTLGAKTAPKACVSOCAAFENPKGLIHGSKARDNN	600
QY	825	DTEGKRYLGHVHNSREPTSEMESSELDQVLOMTFVYSKQOSFALPSNPGNADEECAT	884
Db	601	DTEGKRYLGHVHNSREPTSEMESSELDQVLOMTFVYSKQOSFALPSNPGNADEECAT	660
QY	885	FSAHSGSLKKQSPKATPFCEQOEKEENQGNESNIKEVQVTNITAGFPVYQOKRPVDMNAC	944
Db	661	FSAHSGSLKKQSPKATPFCEQOEKEENQGNESNIKEVQVTNITAGFPVYQOKRPVDMNAC	720
QY	945	SIKGSGRCLSSQOFGNETGLITPKMHGLOPNYKIPLEPIKSVYTKCKKNLLBENE	1004
Db	721	SIKGSGRCLSSQOFGNETGLITPKMHGLOPNYKIPLEPIKSVYTKCKKNLLBENSE	780
QY	1005	EHSMPEREMGENEIPSTVYSTRNNIRBNYKEXASSININEVGSTNEVGSSINENIGSS	1064
Db	781	EHSMPEREMGENEIPSTVYSTRNNIRBNYKEXASSININEVGSTNEVGSSINENIGSS	840
QY	1065	DENIOAELGRNRPKLNMLRLGLVLOPEVYKOSLPGSNCKHPEIRIKOEYEEVQVTNDFE	1124
Db	841	DENIOAELGRNRPKLNMLRLGLVLOPEVYKOSLPGSNCKHPEIRIKOEYEEVQVTNDFE	900
QY	1125	SPYLISDLNEOPMGSSHAQVCSSEFPDDLLDDGETKEDTSPREANDIKESANFSKSVQK	1184
Db	901	SPYLISDLNEOPMGSSHAQVCSSEFPDDLLDDGETKEDTSPREANDIKESANFSKSVQK	960
QY	1185	ELSRSPSPFTHTHLAGYRGAKKLESSEBENLSSDEELPCQOHLFGVNNIPQOSGRH	1244
Db	961	ELSRSPSPFTHTHLAGYRGAKKLESSEBENLSSDEELPCQOHLFGVNNIPQOSGRH	1020
QY	1245	STVATECLSKNTENLULSKNSLNDCSNOVILIAKASQEHNLSEETKCSASLFSQSCSELE	1304
Db	1021	STVATECLSKNTENLULSKNSLNDCSNOVILIAKASQEHNLSEETKCSASLFSQSCSELE	1080
QY	1305	DLTANTNPQDPLIGSSKOMRHQSSQGVGLSDKFLVSDDERGRGLLENNQDEQMSMN	1364
Db	1081	DLTANTNPQDPLIGSSKOMRHQSSQGVGLSDKFLVSDDERGRGLLENNQDEQMSMN	1140

QY	1365 L 1365	
DB	1141 L 1141	
RESULT 6		
ID	046487	PRELIMINARY; PRT: 1140 AA.
AC	046487: 028525:	
DT	01-JUN-1998 (TrEMBLrel. 06, Created)	
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	Breast cancer type 1 susceptibility protein (Breast and ovarian cancer susceptibility protein) (Fragment).	
DE	BRCA1.	
GN	Macaca mulatta (Rhesus macaque).	
OS	Macaca mulatta (Rhesus macaque).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;	
OC	Cercopithecinae; Macaca.	
OX	NCBI_TaxID=9544;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=98122577; PubMed=9462745;	
RA	Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,	
RA	Fodor S.P.A., Brody L.C., Collins F.S.;	
RT	"Evolutionary sequence comparisons using high-density oligonucleotide arrays.";	
RL	Nat. Genet. 18:155-158(1998).	
SC	[2]	
RP	SEQUENCE OF 964-1028 FROM N.A.	
RA	Thompson M.E., Holt J.T.;	
CC	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE	
CC	IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).	
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).	
DR	EMBL: AF019078; AAC39586.1; -	
DR	EMBL: U44730; AAB03212.1; -	
DR	InterPro: IPR002378; BRSCancer.	
DR	PRINTS: PR00493; BRSTCANCER.	
KW	DNA-binding; Nuclear protein; Anti-oncogene.	
FT	NON_TER 1	
FT	DOMAIN 276 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).	
FT	DOMAIN 984 1005 LEUCINE-ZIPPER (POTENTIAL).	
FT	CONFLICT 978 978 R -> Q (IN REF. 2).	
FT	NON_TER 1140 1140	
SO	SEQUENCE 1140 AA; 126958 MW; EC1873CAFEA5E32E7 CRC64;	
Query Match 56.4%; Score 5443.5; DB 6; Length 1140;		
Best Local Similarity 92.6%; Pred. No. 5.4e-273;		
Matches 1057; Conservative 35; Mismatches 47; Indels 3; Gaps 3		
QY	225 ACEPSETDTVNTTEHHOPSNNDLNTTEKRAERHPKTYGGSSVSNLHVEPCGTHTASSIQ 284	
DB	1 ACEFSEKIDTTEHHOSSNNDLNTTEKRAERHPKTYGGSSVSNLHVEPCGTHTASSIQ 60	
QY	285 HEMSSILLTKDRNNVEKAEFCNKSQOPGLARSQHNRMAGSKFQCNDNRPTSEKKVVDLA 344	
DB	61 HEN-SLLTKDRNNVEKAEFCNKSQOPGLARSQHNRMAGSKFQCNDNRPTSEKKVVDLA 119	
QY	345 DPCERKENNKQKLPCESENPRDTEVPVITLSSIQVNMFSKSDLLGSDSDHGESE 404	
DB	120 NALYERKENNKQKLPCESENPRDTEVPVITLSSIQVNMFSKSDLLGSDSDHGESE 179	
QY	405 SNAKVADVLDVLENDVETSGSSEKIDLLASDPHEALICKSERVHKSVSBNIEDKIFGKT 464	
DB	180 SNAKVADVLDVLENDVETSGSSEKIDLLASDPHEALICKSERVHKSVSBNIEDKIFGKT 239	
QY	465 YRKASATPNISSHTELLIIGAVTPEPOLIOERPLTNLKKRRPTSGLPHEDFIKKADLA 524	
DB	240 YRKADPNISSHTELLIIGAVTPEPOLIOERPLTNLKKRRPTSGLPHEDFIKKADLA 299	
QY	525 VQKTPMINIGTNOTEQNGVNNITNSGCHENKTGDSIQNEKNPNPIESLEKSAEKTKA 584	

Db 300 VOKTPEIINOQTNOQONVNTNSAHENKTKGDSIONEKNPNPIESIEESAFKTKA 359  
 QY 585 EPISSISNMELNLINHSKAPKKNLRKRSSTRHHALELVYSRNLSPNCELQIDSC 644  
 Db 360 EPISSISNMELNLINHSKAPKKNLRKRSSTRHHALELVYSRNLSPNCELQIDSC 419  
 QY 645 SSSEIKKKKKYNOMPVRHSRNLQLMGKGPATGAKSKNPNQOTSRRHSDTFPELKLTN 704  
 Db 420 SSSEIKKKKKYNOMPVRHSRNLQLMGKGPATGAKSKNPNQOTSRRHSDTFPELKLTK 479  
 QY 705 AGCSFTKCSNTSELKFEVNPSPREKEKLETVKVSNNADPKDMLSGERYLOTERSV 764  
 Db 480 VQSFNCSNTSE-KEFVNPSLSREKEKELETVKVSNNADPKDMLSGERYLOTERSV 538  
 QY 765 ESSSISLVPGTDYGTQOESISLLEVSTLGAKTEPNKVCQCAAFENPKGLIHGCSNDNN 824  
 Db 539 ESSSISLVPGTDYGTQOESISLLEVSTLGAKTERNKCMQCAAFENPKGLIHGCSNDNN 598  
 QY 825 DTEGFKYPLGHEVNHRSRTSIEMSESLDAQYLQNTFKYSKQSFALFSPNGAEECAT 884  
 Db 599 DTEGFKYPLGHEVNHRSRTSIEMSESLDAQYLQNTFKYSKQSFALFSPNGAEECAT 658  
 QY 885 FSAHSGSLKQSPKVTPECEQKEENOGKABSNIKPVQVNTAGFPVYQOKDPVDNAC 944  
 Db 659 FSAHSGSLKQSPKVTPECEQKEENOGKABSNIKPVQVNTAGFPVYQOKDPVDNAC 718  
 QY 945 STKGSRFLSSQFRNENGLITPNKHGLQNPYRPLPKSPFKTKCKNMLEENR 1004  
 Db 719 STKGSRFLSSQFRNENGLITPNKHGLQNPYRPLPKSPFKTKCKNMLEENR 778  
 QY 1005 EHSMSPEREMNGEN-IPSTVSTISRNINRENVFEKASSNINEVSGSTNEVSGSINIGS 1063  
 Db 779 EHSVSPERAVGNKNIIPSTVSTISHNNIRENNAFKEASSNINEVSGSTNEVSGSINIGS 838  
 QY 1064 SPENIOAELGRNRGRLNMLRGLVQPEVYKOSLPGSNCKPEKTKOYEVEVQTVND 1123  
 Db 839 SPENIOAELGRNRGRLNMLRGLVQPEVYKOSLPGSNCKPEKTKOYEVEVQTVND 898  
 QY 1124 FSPYILSDMLEQPMGSSHAQVCSSETPDDLDDGELKEPTSFANDIKSSAVFSKSYOK 1183  
 Db 899 FSPYILSDMLEQPMGSSHAQVCSSETPDDLDDGELKEPTSFANDIKSSAVFSKSYOK 958  
 QY 1184 GELSRSPFTHTHLAQYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNTIPSQSTR 1243  
 Db 959 GELSRSPFTHTHLAQYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNTIPSQSTR 1018  
 QY 1244 HSTVATECCKNTBEULSLKNSLDCSNQVILAKASQEHHLSEETKCSGSLFSSQCSL 1303  
 Db 1019 HSTVATECCKNTBEULSLKNSLDCSNQVILAKASQEHHLSEETKCSGSLFSSQCSL 1078  
 QY 1304 EDTANTANTQDFPLIGSSKOMRHOSESQVGLSDKELVSDDERGGLGLENNOEBSMS 1363  
 Db 1079 EDTANTANTQDFPLIGSSKOMRHOSESQVGLSDKELVSDDERGGLGLENNOEBSMS 1138  
 QY 1364 NL 1365  
 Db 1139 NL 1140  
 RESULT 7  
 046488 PRELIMINARY: PRT: 1146 AA.  
 AC 046488;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE Breast cancer type I susceptibility protein (Breast and ovarian cancer susceptibility protein) (Fragment).  
 GN BRCAL.  
 OS Alouatta seniculus (Howler monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Alouattinae; Alouatta.

OX NCBI\_TaxID=9503;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98122577; PubMed-9462745;  
 RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M., Fodor S.P.A., Brody L.C., Collins F.S.;  
 RT "Evolutionary sequence comparisons using high-density oligonucleotide arrays";  
 RL Nat. Genet. 18:155-156(1998).  
 CC -1- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 DR EMBL: AF019079; AAC39587.1; -;  
 DR InterPro: IPR002378; Brst.cancer1.  
 DR PRINTS: PR00493; BRSTCANCERT.  
 KW DNA-binding; Nuclear protein; Anti-oncogene.  
 FT NON\_TER 1  
 FT DOMAIN 278 285 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 200 203 POLY-SER.  
 FT DOMAIN 815 818 POLY-SER.  
 FT DOMAIN 1001 1004 POLY-GLU.  
 FT DOMAIN 996 1005 LEUCINE-ZIPPER (POTENTIAL).  
 FT NON\_TER 1146  
 SQ SEQUENCE 1146 AA; 128119 MW; BFD07EBBF851890E CRC64;  
 Query Match 52.0%; Score 5014.5; DB 6; Length 1146;  
 Best Local Similarity 86.2%; Pred. No. 7.9e-251;  
 Matches 991; Conservative 55; Mismatches 90; Indels 13; Gaps 7;  
 QY 226 CERSEEDVNTNTEHQPNSNDLNTTEKRAARHEPEKYQGSVSNLHVEPGTNTNASSLOH 285  
 Db 2 CERSEEDVNTNTEHQPNSNDLNTTEKRAARHEPEKYQGSVSNLHVEPGTNTNASSLOH 61  
 QY 286 ENSLLTLTKDRMVEAEFCNKSQOPGLARSOHNRNAGSKETCNDRPTSTEEKVDLND 345  
 Db 62 ENSLLTLTKDRMVEAEFCNKSQOPGLARSOHNRNAGSKETCNDRPTSTEEKVDLND 121  
 QY 346 PLCKEREMKOKPLPCSENPDR-DTEEDVPWITLNSIOKVNMFPSPRELLGSDSHGGESE 404  
 Db 122 PLHGRKEMKOKPCCGSENPDRDTEEDVAMILNLSIOKVNMFPSPRELLGSDSHGGESE 181  
 QY 405 SNAKVADVILNENDEYSGSSSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGRT 464  
 Db 182 SNAKVADVILNENDEYSGSSSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGRT 241  
 QY 465 YKKKASLPMLSHVTEMLIIGAVTEPOIIQERPLTKIKRRRPTSGLHPDEFIKKADLA 524  
 Db 242 YKKKASLPMLSHVTEMLIIGAVTEPOIIQERPLTKIKRRRPTSGLHPDEFIKKADLA 301  
 QY 525 VOKTPEIINOQTNOQONVNTNSAHENKTKGDSIONEKNPNPIESIEESAFKTKA 584  
 Db 302 VOKTPEIINOQTNOQONVNTNSAHENKTKGDSIONEKNPNPIESIEESAFKTKA 360  
 QY 585 EPISSISNMELNLINHSKAPKKNLRKRSSTRHHALELVYSRNLSPNCELQIDSC 644  
 Db 361 EPISSISNMELNLINHSKAPKKNLRKRSSTRHHALELVYSRNLSPNCELQIDSC 420  
 QY 645 SSSEIKKKKKYNOMPVRHSRNLQLMGKGPATGAKSKNPNQOTSRRHSDTFPELKLTN 704  
 Db 421 SSSEIKKKKKYNOMPVRHSRNLQLMGKGPATGAKSKNPNQOTSRRHSDTFPELKLTN 480  
 QY 705 AGCSFTKCSNTSELKFEVNPSPREKEKLETVKVSNNADPKDMLSGERYLOTERSV 764  
 Db 481 AGCSFTKCSNTSELKFEVNPSPREKEKLETVKVSNNADPKDMLSGERYLOTERSV 540  
 QY 765 ESSSISLVPGTDYGTQOESISLLEVSTLGAKTEPNKVCQCAAFENPKGLIHGCSNDNN 824  
 Db 541 ESSSISLVPGTDYGTQOESISLLEVSTLGAKTEPNKVCQCAAFENPKGLIHGCSNDNN 600  
 QY 825 DTEGFKYPLGHEVNHRSRTSIEMSESLDAQYLQNTFKYSKQSFALFSPNGAEECAT 884  
 Db 601 DTEGFKYPLGHEVNHRSRTSIEMSESLDAQYLQNTFKYSKQSFALFSPNGAEECAT 660

ID	Q9%ID2	PRELIMINARY;	PRT;	1812 AA.
QY	885	FSASGSLKTKKSPKVTPECECKEENOCKNSN1KPVQTVITAGFPVYQKDKPVDNAK	944	
Dd	661	FSACSRSLKTKKSPKVTPECECKEENOCKESKES1EETVVT7TGFPPVYCKDKPVDYAR	720	
QY	945	S1KSGSRFCLSSQSRGNETGLTIPNKGGLQNPYRIPPLPIKSFVYTKCKKMLLEENF	1004	
Dd	721	-IEGSRCLSSQSRGNETGLTIPNKGGLQNPYHMSPLIPTRSPVYTKCKKMLLEENF	779	
QY	1005	EHMSSPEREMGEN-1STVSTISRRNNIRENVFREASSNI-----NEVGSSTNEVGS	1056	
Dd	780	EHSMSPRAMGNKNI1STVSTISHNN-RENAFKEITSSSIYEVGSSSTNEAGSSTNEVGS	838	
QY	1057	S1NIGSSDENIQAELGRNRPKNALRLGVLOPEYVKSLPGNSCKHPKIKOEYEEY	1116	
Dd	839	S1NEVGSSDENIQAELGRNRPKNALRLGVLOPELCKSLTISDCKHKEIKOEHEV	898	
QY	1117	VQVYNTDPSPLISDNLEQPMGSSHASQVSETPDDLLDGEIKEDTSFANDIKSSAV	1176	
Dd	899	VQVYNTDPSPLISDNLEQPMGSSHASQVSETPDDLLDGEIKEDTSFANDIKSTV	958	
QY	1177	FSKSVQKGLSRSPFTHTHLAGYRRGAKKLESSEENLSDPELPCFOHLLFKVNN	1236	
Dd	959	FSKSVQKGLSRSPFTHTHLAGYRRGAKKLESSEENLSDPELPCFOHLLFKVNS	1018	
QY	1237	IPSGSTHSTVATECLSKNTEENILSLKNSLDCSNQVILLAKASQEHHLSEFTKCSALF	1296	
Dd	1019	TP-OSTRSTSAECLSKNTEENILSLKNSLDCSNQVILLAKASQEHHLSEFTKCSALF	1077	
QY	1297	SSQSELEDLTANNTPDPLIGSSKMMRROSQVGLSDKELVSDDERGTGLEENNO	1356	
Dd	1078	SSQSELEDLTANNTPDPLIGSSKMMRROSQVGLSDKELVSDDERGTGLEENNO	1137	
QY	1357	EEQSMDSNL 1365		
Dd	1138	EEQSMDSNL 1146		

[illegible]

```

Db      942 RSENGISATGKSGISQNSHFQKQSVSPIRRSIKTDNRKPLTEGRFERHTSTTEMAVGNEN 1001
Qy      1019 I-PSYSTISRNRIREVVEFEASSNINEVGSSINIGSSDEDTQAEELGRNG 1077
Db      1002 ILQSTVHTVSLN-RGNACEAGS-----GSIHEVCSTGGSFPQGLNRNG 1046
Qy      1078 PKINAMLRGLVLOPEYVKOSLPGSNCKHPKIKOEYEVVQVNTFSPFLIDNLEQPM 1137
Db      1047 PKVNTVPLDLSMOPGCGQGVSPVD-KYLEIKQEGEAVC----ADSPCLPFDHLEQSM 1101
Qy      1138 GSSHASQVCEPDDLLDQGEIKEDTSFAENDIKESAVFSKSVQKGLSRSPPTHTH 1197
Db      1102 -SGKVFQVCEPDDLLDQGEIKEDTSFAENDIKESAVFSKSVQKGLSRSPPTHTH 1160
Qy      1198 LAGYIRGAKKLSSSENLSEDELPQCHLLFGKVNRIPSQSTRSHVATCLSKNTE 1257
Db      1161 KQSLHRASRKLSSESDSTEDDLPQCHLL-SRISNTP-ELTRCRSAVVTGIEKAE 1218
Qy      1258 ENLLSKNSLNDSCNOVILAKAQEHLSSETKCSAFLFSQCELEDTLANTNTODPFL 1317
Db      1219 GTAPAKKGGSSDCNNVEIMTEASQEHQFSDPRCSGRMFSQNSAAGSTANNSQDSNF 1278
Qy      1318 IGSSKQMRHQSESQGVGLSDKELVSDDEERGTLLENNOEQSMDSNLGRA-ASGCESET 1376
Db      1279 IPPSKQRSHQCGNEAFSLDKELISDNEWATCLEEDNDQEE--DIIIPDEASAGESET 1336
Qy      1377 SVSEDCSGLSQSDIITTOQRDMONHLIKLOEAMLEVLVQKHQSQPSNSTPSTISDS 1436
Db      1337 NLSIEDC-----SOSDILITTOQRATMKYNLKLQOEMAHLEVLVQKHQSQPSNSTPSTISADP 1392
Qy      1437 SALEDLRNPEQSTSEKAVLTQSKSEYPISONEGLSADFEYS-ADSSPSKKEGVER 1495
Db      1393 CALEDPLDLEPNMNGAAILTSKINENPVQNLKACDDKFOQLHLEGPISGDESGMGR 1452
Qy      1496 SPSKCPSLDDRWYMHSCSGSLQNRNYPQSEELIKVYDVEEQULESGPHDLTETSYLPR 1555
Db      1453 PSEPKSPLAGSRGSAHGCSRHLLQKRNPSQOELLQAGSE----ASSEPHNSTGQCLPR 1508
Qy      1556 QDEGPPYLESGLISLSD-DPESDPEDRPRPESARNGNIPSSALKVQKLAENASQP 1614
Db      1509 RELGEPYLGSGISLSSRDPESESPK---EPANHGTTPASTSAKIPGQVAFPSAA 1564
Qy      1615 AAHTTIDTGYNMEESVSEKPELTASTERVNRKSMNVVSGTLPEEFMLVYFARKHHT 1674
Db      1565 AGAD-----KAVGIVSKIKELTSSERADROISMVYSGTLPEEFMLVYFARKHHT 1617
Qy      1675 TLTNLTETTTVHVKMTDAEFCERTLKYYFLGIAGKGVVSYFWYQSIKERMLNHDH 1734
Db      1618 TLTDAITEETTHVYIKTDAEFVCEFTLKYYFLGIAGKGVVSYFWYQSIKERMLNHDH 1677
Qy      1735 EVAGDVVNRNNGPKRARESDOKRIFRGLICTCYPTNMPDQLEMMVQGLGASVKE 1794
Db      1678 EVKGDVAVGNNGPPRRSRSRE-KLEKGLQVYCCPFTNMPKDELEMLQGLGASVKE 1736
Qy      1795 LSSFTLTGTGVHPIVWVQPDAMTENGFMHAIQGCCEAPVYTRVWDLSEVALYQCELDITY 1854
Db      1737 LPSLTHTDGAHLVIVQPSAMTSDNSCPDGLCKARLVAMVWDLSSLYRCDLDAYL 1796
Qy      1855 IPQI 1858
Db      1797 VQNI 1800

```

```

RESULT 9
ID 054952 PRELIMINARY; PRT; 1817 AA.
AC 054952; P97951;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Breast cancer type I susceptibility protein (Breast and ovarian cancer
susceptibility protein).

```

```

GN BRCA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=9911388; PubMed=9892727;
RA Bennett L.M., Brownlee H.A., Hagavik S., Wiseman R.W.;
RT "Sequence analysis of the rat brca1 homolog and its promoter region.";
RN Mamm. Genome 10:19-25(1999).
[2]
RP SEQUENCE OF 8-222 FROM N.A.
RC STRAIN=WISTAR-KYOTO; TISSUE=SPLEEN;
RX MEDLINE=96358532; PubMed=8761410;
RA Chen K.S., Shepel L.A., Haag J.D., Heil G.M., Gould M.N.;
RT "Cloning, genetic mapping and expression studies of the rat Brca1
gene.";
RT Carcinogenesis 17:1561-1566(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION, MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL: AF036760; AAC36493.1; -.
DR EMBL: S82504; NOT_ANNOTATED_CDS.
DR EMBL: S82502; NOT_ANNOTATED_CDS.
DR EMBL: U060523; AAB40387.1; -.
DR EMBL: S82500; AAB37501.1; -.
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR002378; Brst_cancerl.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00533; BRCT_2.
DR Pfam: PF00097; Zf-C3HC4_1.
DR PRINTS: PR00493; BRSTCANCER1.
DR SMART: SM00292; BRCT; 2.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS50172; BRCT; 2.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene.
FT ZN_FING 24 64
FT DOMAIN 497 503 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 38 38 Q -> K (IN REF. 2).
FT CONFLICT 192 192 A -> M (IN REF. 2).
SQ SEQUENCE 1817 AA; 199876 MW; COB4760F0E349A01 CMC64;

Query Match 50.5%; Score 4873.5; DB 11; Length 1817;
Best Local Similarity 56.2%; Pred. No. 2,9e-243;
Matches 1049; Conservative 259; Mismatches 488; Indels 69; Gaps 32;

```



```

Db      1 AEEFSE-DITNTEHNSCKDLNTIEKATERPEDEYSSVSNLHVPCGTHTHASSIQ 59
OY      285 HENSLLTKDRMNVKAEFCNKSOPGLARSOHNWAGSKENCNDRPSTTEKKYDLNA 344
Db      60 HENSLLTKDRMNVKAEFCNKSOPGLARSOESWAGSKENCNDRPSTTEKKYDLNA 119
OY      345 DPLCEKERNKOKLPCSENPRDTEDEVWITLNSIQKVNEMFSRDELSDSHGES 404
Db      120 DSHGKERNKOPPRPESPROTQDVPWITLNSIQKVNEMFSRDELSDSHGES 179
OY      405 SNAKVAIDLVLNVEDEYSGSSEKIDLASDPHEALICKSERVHKSVESNIEDKIFGKT 464
Db      180 SNAKVAIDLEVPNDVGYDSSEKIDLASDPHEALICKSERHSPVESNIEDKIFGKT 239
OY      465 YRRKASLPLNLSHYTEMLIIGAFVTEPOITIOERPLTKLKRKRRTTSGLHPEDIKKADLA 524
Db      240 YOKRTSLPLNLSHYTEMLIIGAFVTEPOITIOERPLTKLKRKRRTTSGLHPEDIKKADLA 299
OY      525 -YOKTPEMINOGTNOFGOVANNITNSGHENKTKGDSIQNEKNPPIESLEKESAFKTK 583
Db      300 YOKTPEKINOGIDQEQNDVANNITNSGHENKTDQYQKKNANPTESLEKESAFKTK 359
OY      584 AEPISISSIMLELNIHNSKAPKRNLRKRSSTRIHLELVYSRNLSPPNCTELQIDS 643
Db      360 AEPISISSIMLELNIHNSKSKRNLRKRSSTRIHLELVYSRNLSPPNCTELQIDS 419
OY      644 CSSESEIKKKYNNOMVRSRNLQMEKREPATGAKSKSNPNOTSRTSDOFFELKTX 703
Db      420 CSSESEIEKKNSSQVVRHSRKLQLENEKREPATGAKSKSNPNQISRRHSNAFPELRILT 479
OY      704 NAPSFTKCSNTESELKEFVNPSLPREKEELETETVVSNNADPKDLMSEGERVLOTERS 763
Db      480 NYVFEFANSSSKKLOEFIDPQLQREIEENLETIHVNSAKRPDLVLSGEKGLQTERS 539
OY      764 VESSISILVPGDYGQESISLLEVSTLCKAKTEPNKCVSQCAAFNPGLIHGSKDNR 823
Db      540 VESTISILVPGDYGQESISLLEVSTLCKAKTAPQAHNOCAIINPEKELIHGSKDNR 599
OY      824 NDEGEFKYPLGHEVNSRFTSIEMESSELDAQVLOMTFVSKRSKSFALSPNNAEECA 883
Db      600 NDEGEFKYPLGHEVNSRFTSIEMESSELDAQVLOMTFVSKRSKSFALSPNNAEECA 659
OY      884 TFSAHSGSLKOKSPKTFECCQKEENOGKSNIRKQVOTVNTAGPPVVGOK -DKPVDNA 942
Db      660 TYVAHRLRLKOKSPKTFECCQKEENOGKSNIRKQVOTVNTAGPPVVGOK -DKPVDNA 719
OY      943 KCSIKGSRFCUSSQFRGNETGLTPNKHGLLONPYRIPPLPIKSFVKTKCKNLLEN 1002
Db      720 KESIKGSRFCUSSQFRGNETGLTPNKHGLLONPYRIPPLPIKSFVKTKCKNLLEN 779
OY      1003 FEHSHSPREKGNEN-IPSTVTSIRNNIRENVFKEASSNINENEGSSITNEVGSSINEL 1061
Db      780 FEHSHSPREKGNEN-IPSTVTSIRNNIRENVFKEASSNINENEGSSITNEVGSSINEL 839
OY      1062 GSSDENIQALGRNRPKLNAMLRGLVLOPEVYKOSLPPSNCKHPRIKQOEVEVYQVNA 1121
Db      840 GSGENIQALGRNRPKLNAMLRGLVLOPEVYKOSLPPSNCKHPRIKQOEVEVYQVNA 899
OY      1122 TPFSPILSDNLEQPWGSHASQVCEPDDLLDGEIKEDTSPFAENDIKESSAVFSKV 1181
Db      900 TMLSTCLLSHNEQPESSHASQVCEPDDLLDGEIKEDTSPFAENDIKESSAVFSKV 959
OY      1182 OKGELSRSPPTPHHLLQGRGAKKLESSEENLSEDEELPCPHLLFGVYNNIPSS 1241
Db      960 OKGELSRSPPTPHHLLQGRGAKKLESSEENLSEDEELPCPHLLFGVYNNIPSS 1019
OY      1242 TRHSTVATECLSKNTEENLSTKNSLNDCSNOVYILAKAQEHLSSETKCSALFSSQCS 1301
Db      1020 TRHSTVATECLSKNTEENLSTKNSLNDCSNOVYILAKAQEHLSSETKCSALFSSQCS 1079
OY      1302 ELEDLTANTNTODPFLI--GSSKQMRHQSQGVGSLDKELVSDDERGTGLEE-NNQEE 1358

```

```

Db      1080 AVELATNTNTODPFLIMFPPSNHVRHSENOEVLSDKELVSDDERETALEYNNHEE 1139
OY      1359 OSMSDNL 1365
Db      1140 OSVDSNL 1146

RESULT 11
046489
ID 046489 PRELIMINARY; PRT: 1141 AA.
AC 046489;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
susceptibility protein) (Fragment).
GN BRCA1.
OS Galago crassicaudatus (Thick-tailed galago) (Otolemur crassicaudatus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otolemur.
OX NCBI_TaxID=9463;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
arrays."
RL Nat. Genet. 18:155-158(1998).
CC -i- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -i- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL: AF019080; AAC39588.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON-TER 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 199 202 POLY-SER.
FT DOMAIN 816 819 POLY-SER.
FT NON-TER 1141 1141
SQ SEQUENCE 1141 AA; 127578 MW; 2653CA894834946F CRC64;

Query Match 41.9%; Score 4040; DB 6; Length 1141;
Best Local Similarity 72.18; Pred. No. 1.7e-200;
Matches 831; Conservative 105; Mismatches 192; Indels 24; Gaps 15;

OY      226 CEFSELDVNTNTEHNSCKDLNTIEKATERPEDEYSSVSNLHVPCGTHTHASSIQ 285
Db      2 CEFSELDVNTNTEHNSCKDLNTIEKATERPEDEYSSVSNLHVPCGTHTHASSIQ 61
OY      286 HENSLLTKDRMNVKAEFCNKSOPGLARSOHNWAGSKENCNDRPSTTEKKYDLNA 345
Db      62 HENSLLTKDRMNVKAEFCNKSOPGLARSOESWAGSKENCNDRPSTTEKKYDLNA 121
OY      346 DPLCEKERNKOKLPCSENPRDTEDEVWITLNSIQKVNEMFSRDELSDSHGES 405
Db      122 DPLCEKERNKOPPRPESPROTQDVPWITLNSIQKVNEMFSRDELSDSHGES 181
OY      406 SNAKVAIDLVLNVEDEYSGSSEKIDLASDPHEALICKSERVHKSVESNIEDKIFGKT 465
Db      182 SNAKVAIDLEVPNDVGYDSSEKIDLASDPHEALICKSERHSPVESNIEDKIFGKT 241
OY      466 YRRKASLPLNLSHYTEMLIIGAFVTEPOITIOERPLTKLKRKRRTTSGLHPEDIKKADLA 524
Db      242 YOKRTSLPLNLSHYTEMLIIGAFVTEPOITIOERPLTKLKRKRRTTSGLHPEDIKKADLA 301
OY      525 -YOKTPEMINOGTNOFGOVANNITNSGHENKTKGDSIQNEKNPPIESLEKESAFKTK 584
Db      302 YOKTPEKINOGIDQEQNDVANNITNSGHENKTDQYQKKNANPTESLEKESAFKTK 361
OY      585 AEPISISSIMLELNIHNSKAPKRNLRKRSSTRIHLELVYSRNLSPPNCTELQIDS 644

```



```

Db 362 EPISSISNMKLELNIHNSKASKKKRRKSSRHIRALELVNKNPSPHNTLQJOSC 421
OY 645 SSSEIKKKKYNOMPVHRSHRNQJLMEGKEPATGAKSKNPNQTSKRHSDPELKLKN 704
Db 422 SSSEIKDKSSDQJFVRHSKRKPGLMEDREPATGAKSKNPNQISKRHVSDFPEVALTN 481
OY 705 APGSFTKCSNTESEKFEVNPSPILPREKEEKL-ETVKSNNADPKDMLSGERVLO--T 760
Db 482 ISSFTNCSGSRNLKEFVNPSPILQKRTTEENLEETIOVSNTSGP---VLGSRVLIQISE 538
OY 761 ERSVSSSSISLVPGDYTOESISLLEVTSTLGAKTERPKVCSCAFENPGLHGCCK 820
Db 539 ERSIKSTISILVPTDYTOESISLLEVTSTLGAKTERPKVCSCAFENPGLHGCCK 598
OY 821 DNRNDTEGFKYPLGHEVNSRSTSIEMESSELDQYLONTFVKRSKOSFALSNPGMAE 880
Db 599 DTGNDTEGKPLRHEINHIOKISMEDSELDYLONTFVKRSKOSFALSNPGMAE 655
OY 881 ECATFSAH--SGSLKKQSPKYTECEQKEENQKNESNIKPVQYVITAGFPVQOKDP 938
Db 656 ECATFCAOSLSASLRKGS--KYLCEQ--IENPMKEPKIKHIGNNINMGFSVQOKDK 713
OY 939 VDN-AKCSIKGSRFCLSSQFRGNETGLITPKHGLLONPRIPPLPKSVTKCKKN 997
Db 714 TDDYAKSIAKASRCLSNQFRDNETESITVKKLGLIOLNLYHIPPLSPFLDKTKMTN 773
OY 998 LLENFEEHSMSPEREMNEN-IPSTVSTISRNINRENVFKASSNINEVSGSTINEVGS 1056
Db 774 LLEERFERHSHVLPKAVGENTVPSTMTINONN-RESAYKRASSSSINEVSGSTINEVGS 832
OY 1057 SINEIGSSDENIQALGNRGPKNLMLRLGLVQPEVYKQSLPGSNCKHPEIKQIEEY 1116
Db 833 SVNEVGSSSENIQALGNRGPKNLMLRLGLVQPEVYKQSLPGSNCKHPEIKQIEEY 892
OY 1117 VQVNTPEFVPLIDONLEOPGSSHAQVSCETPDLLDGEIKQDTSAAEDIESSAV 1176
Db 893 VQVNPDSCLISNLEQPTRSSHASQJLSETPDLLDDELKENTSAENNIERSAV 952
OY 1177 FSKSVQKGLSRSPFTHTLHAQGRGAKKLESSEENLSEDELEPCFQHLFGKYN 1236
Db 953 FSKVMRREIRSPSPPLAHILHQAQREVRKLESSEENLSEDELEPCFQHLFGKYN 1012
OY 1237 IPSOSTRSTVATECLNTEENLILKNSLNDSCNOYIILAKASOEHLSETKCSAIF 1296
Db 1013 TSSGSTSTRTATCLSRKTEETLVSLONTLNDSCNOYIILAKASOEHLSETKCSAIF 1072
OY 1297 SSQCSLELDTANTTOPFLIGS--SKOMRHQSSQGVGLSDKELVSDDEERGGLLE- 1353
Db 1073 SSQCS--EDLTANTNTQDFILMSDPLSKOMRQDSQENQY-LSDNELVSNDEDEREPLSD 1129
OY 1354 NNOEQGMSDNL 1365
Db 1130 NHOEQSVSDNL 1141

```

```

RESULT 12
O924EO PRELIMINARY: PRT: 1063 AA.
AC O924EO:
DT 01-DEC-2001 (Tremblrel. 19. Created)
DT 01-DEC-2001 (Tremblrel. 19. Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19. Last annotation update)
DE BRCA1 (Fragment).
GN BRCA1.
OS Pedetes capensis (Springhaas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Pedetidae; Pedetes.
OX NCBI_TaxID=10023;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21219190; PubMed=11319262;
RA Adkins R.M., Gelke E.L., Rowe D., Honeycutt R.L.;
RT "Molecular phylogeny and divergence time estimates for major rodent

```

```

RT groups: evidence from multiple genes.;
RL Mol. Biol. Evol. 18:777-791(2001).
DR EMBL: AF332047; AAK71627.1; -.
FT NON-TER 1
FT NON-TER 1
SQ SEQUENCE 1063 AA; 118726 MW; 9B9CE78FE404745 CRC64;
Query Match
38.3%; Score 3693; DB 11; Length 1063;
Best Local Similarity 69.5%; Pred. No. 1,3e-182;
Matches 752; Conservative 103; Mismatches 203; Indels 24; Gaps 7;
OY 262 GGSVSNLHVEPCGNTTHASSLQEHNSLLTKDRMNYEKAFCNKSQOPCLARSQHNW 321
Db 1 QGISVSNLCEPBCGNTTHASSLQEHNSLLTKDRMNYEKAFCNKSQOPCLARSQHNW 60
OY 322 AGSKETCNDRTPTSEKRYVDLADNADLCGRKEMNKOKLCCSNPHTDEVPIITLNSSTOK 381
Db 61 AASKETCNDRTPTSEKRYVDLADNADLCGRKEMNKOKLCCSNPHTDEVPIITLNSSTOK 120
OY 382 VNEWFSRDELLGSDSHDGESESNAKVADYLDVLENYDEYSGSSEKIDLLASDPHEALI 441
Db 121 VNEWFSRDEMLTSDDLHDGVSNAEYLGALVDELDKHSQSEKIDLLASDPHEALIT 180
OY 442 CKSERVHSKVSNEIDKIFGKTYRKASLPNLSHVTENLITGAFVTEPQIIOERPLTNK 501
Db 181 CKSERVCAKPVESNIEDKVFCKTYRKISLPNLSPITENLITGAFVTEPQIIOERPLTNK 240
OY 502 LKRRRPRSGHPEPFKIKADLAV-QKPEMINQNTOTEOGQVMTNTEGCHENKRGD 560
Db 241 LKRRKRAVSPLEPFIKVDLVLOKTPKINQTAQMGQNGOVINNTNLDLENKND 300
OY 561 STONEKNPPIESLESSEAFYTKAPISSTISNMELELNHNSKAPKKNRLRRKSSTRI 620
Db 301 YLKEKNANPIESLESSEHTKAPISSTISNMELELNHNSKAPKKNRLRRKSSTRI 360
OY 621 HALELVSRNLSPPNCTELQIDSCSSSEIKKKNQMPVHRSHRNQJLMEGKEPATGAKK 680
Db 361 HALELVVKNKSPNPQTELOIDSCSSSNEIKKCSQDSVSRKQJLMEGKEPATGAKK 420
OY 681 SNKPNEQTSKRHSDTPFELKLTNAPGSFTKCSNTESEKFEVNPSPILPREKEEKLTV 740
Db 421 NNTPEKTSKRHSDTPFELKLTNAPGSFTKCSNTESEKFEVNPSPILPREKEEKLTV 480
OY 741 SNNADPKDMLSGERVLOTERSVSSSSISLVPGDYTOESISLLEVTSTLGAKTERPK 800
Db 481 SNNADPKDMLSGERVLOTERSVSSSSISLVPGDYTOESISLLEVTSTLGAKTERPK 540
OY 801 CYSQCAFENPGLHGCCKDNKRNTEBGFKYPPLGHEVNSRSTSIEMESSELDQYLONT 860
Db 541 CMTQYVALEHPRKELSHSCSKDTRNDTKGKDALREVNHIQETMEEESLDQYLONT 600
OY 861 FKVSRQSFALFSGNNAEEECATFSAHSGSLKOSPVPTEFCQKEENQKNESNIKPV 920
Db 601 FOASKRQSFALFPNARNPEKEC----AHSMLRKQSTVTLTEYQKEENQKESKIKPA 656
OY 921 QTVNTTAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRGNETGLITPKHGLLQNPYRI 980
Db 657 QTVNTTAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRGNETGLITPKHGLLQNPYRI 716
OY 981 PPLPIKSPVTKCKKNLLENFEEHSMSPEREMNEN-IPSTVSTISRNINRENVKKA 1039
Db 717 PSISPVKSFITTKYKKSPESEKFEKYSASHNAGNESIIQSIYTTIGQNNIRENAKKA 776
OY 1040 SSSNINEVSGSTINEVSGSSELDQYLONTFVKRSKOSFALSNPGMAE 1099
Db 777 SS-----GSINEVSGSSELDQYLONTFVKRSKOSFALSNPGMAE 822
OY 1100 GSNCKHPEIKK-OEYEEVQTVNTDFSPYLLSDMLDEQPMGSSHAQVSCETPDLLDGE 1158
Db 823 VSS-KHPMKKQOENEGVQAVADAFSPCLLSNSLEQPMRSGCVSQJLSETPDLLDKE 881
OY 1159 IKEDTSFENDIKESSAVFSKSVQKGLSRSPFTHTLHAQGRGAKKLESSEENLSS 1218

```



Db 882 VKENTSTEGGIKETSAVFSKSIQGESSSPSLTHISLAOCHORRARKLESSEISS 941  
 QY 1219 EDEELPCFOHLFGKYNINIPSTOSTRSTVATECLSKNTEENLSLKNLSNDCSNVOYLAK 1278  
 Db 942 EDEELPCFOHLFGKYNINIPSTOSTRSTVATECLSKNTEENLSLKNLSNDCSNVOYLAK 1001  
 QY 1279 ASQEHHLSEETKCSAFLSSQCELEDTANTNTODPFLI--GSSKOMRHOSESGVGLS 1336  
 Db 1002 TSOEPLHSEAKYSGLSSFLSSQCELEDTANTNTODPFLI--GSSKOMRHOSESGVGLS 1061  
 QY 1337 DK 1338  
 Db 1062 DK 1063

RESULT 13  
 Q8MMT3 PRELIMINARY: PRT: 947 AA.  
 ID 08MMT3  
 AC 08MMT3  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE BRCA1 (Fragment).  
 OS Tapirus pinchaque (Mountain tapir).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.  
 OX NCBI\_TaxID=30582;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Murphy W.J., Eizirik E., O'Brien S.J., Madsen O., Scally M., Douady C.,  
 RA Teeling E., Ryder O., Stanhope M.J., deJong W.W., Springer M.S.;  
 RT "Resolution of the early placental mammal radiation using Bayesian  
 RT phylogenetics.";  
 RL Science 0:0-0(2001).  
 DR EMBL; AY057830; AAL30077.1; -  
 FT NON\_TER 1  
 FT NON\_TER 947  
 FT NON\_TER 947  
 SEQUENCE 947 AA; 104734 MW; 21D8DE08EBDE0D43 CRC64;

Query Match 36.8%; Score 3557.5; DB 6; Length 947;  
 Best Local Similarity 75.1%; Pred. No. 1.1e-175;  
 Matches 713; Conservative 79; Mismatches 152; Indels 5; Gaps 5;  
 QY 274 CGTNTHASSLOHENSLLTKDRMNVKAEPCNKSKOPGLARSONHRMAGSKETCNDRT 333  
 Db 1 CGTNTHASSLOHENSLLTKDRMNVKAEPCNKSKOPGLARSONHRMAGSKETCNDRT 60  
 QY 334 PSTERKVDLADPLCEKEMNKOKLPCSENPRTDEVPITLNNSTIOKYNEMFSRSDLL 393  
 Db 61 PSEKVKLVNADPLGKREKLNKQKPCNSNSRSDQDVPITLNNSTIOKYNEMFSRSDLL 120  
 QY 394 GSDSDHDESESNAYADVLNLEVDEYSGSSEKIDILASPHREALICKSRVHSKSYE 453  
 Db 121 SSDSDRSDGPEHTEVAGAVEPNEVDGYSSEKIDILASPPALICESERVHSKPYA 180  
 QY 454 SNIEKITEGTYRKASAPLNSHVTENILGAFVPEPOLIOBRPLTNKLRKRRTSGH 513  
 Db 181 NIEKITEGTYRKASAPLNSHVTENILGAFVPEPOLIOBRPLTNKLRKRRTSGH 240  
 QY 514 PEDFTKKADLA-VOKTPKEMINGTNOTEQNGOVNITNSGHNKTKGDSIQNEKPNPPE 572  
 Db 241 PEDFTKKADLA-VOKTPKEMINGTNOTEQNGOVNITNSGHNKTKGDSIQNEKPNPPE 300  
 QY 573 SLEKSAFTKAEPLISSISNNLELNIHNSKAPKKNRLKRSSTRHIALELVVSRNLS 632  
 Db 301 SLEKSAFTKAEPLISSISNNLELNIHNSKAPKKNRLKRSSTRHIALELVVSRNLS 360  
 QY 633 PNCGELOIDSSSSSEIKKKKKYNNOMPVHNSRLNIMEKEPATGAKKSNKNEQTSKH 692  
 Db 361 PNHTELQIDCCSSSEIKKKKKYNNOMPVHNSRLNIMEKEPATGAKKSNKNEQTSKH 420  
 QY 693 DSDTEPELKLITNAPGSLTKCSNTSELKEFVNPSLPEEREKRLTETVYKSNNAEDPKDML 752

Db 421 ASHAFPELKLITNAPGSLTKCSNTSELKEFVNPSLPEEREKRLTETVYKSNNAEDPKDML 480  
 QY 753 SGERVLQTRSVSESSISLVPGETDYGTQESISLEVESTIGRAKTEPNKCVSCAAFEENK 812  
 Db 481 SGERVLQTRSVSESSISLVPGETDYGTQESISLEVESTIGRAKTEPNKCVSCAAFEENK 540  
 QY 813 GLIHGSKNRNDTEFFKYPGLCHEVNSRETSIEMESELDOYLONTFKVSKROGFALE 872  
 Db 541 ELMHGSCKDTRNDTEFFKYPGLCHEVNSRETSIEMESELDOYLONTFKVSKROGFALE 600  
 QY 873 SNPGNAEECATFSAHSGSLKQSPKVTCEQKEENOGKNSNIRPVQTVNITAGEPVV 932  
 Db 601 SNSGN-PEECATVCAHSPRLKRESKPVLTCEGRKEENOGKNSNIRPVQTVNITAGEPVV 659  
 QY 933 GQKDRPVDAKCSIKGSRFCSSQFRGNETGLTPNKIGLLONPRIRPLPIKSFVAT 992  
 Db 660 GQKDRPVDAKCSIKGSRFCSSQFRGNETGLTPNKIGLLONPRIRPLPIKSFVAT 719  
 QY 993 KCKKNLLEENFEHSSKSPREKNGEN-IPSTVSTIRNNIRENVEKASSNINEVGST 1051  
 Db 720 TCKKNVSEKLEHSSVSPRAKNGESITOSTYSTISQNNIRSTREKVEGSSSINEVGST 779  
 QY 1052 NEVGSSINEGSSDENIOAELRNRCPLNMLRLGLVLOPEVYKOSLPSSCKHPEIKRO 1111  
 Db 780 NEVGSSINEGSSDENIOAELRNRCPLNMLRLGLVLOPEVYKOSLPSSCKHPEIKRO 839  
 QY 1112 -EYEVQVNTDFSYLISDNLQPMGSSHASQVCEFPDLDLDDGELKEDTSFENDI 1170  
 Db 840 GENEQVQAVNTDFSYLISDNLQPMGSSHASQVCEFPDLDLDDGELKEDTSFENDI 898  
 QY 1171 KESSAVFSKSVOKGELSRSPFTHLAOGYRGAKKLESSEENLSE 1219  
 Db 899 KESSAVFSKSVOKGELSRSPFTHLAOGYRGAKKLESSEENLSE 947

RESULT 14  
 Q9BDV3 PRELIMINARY: PRT: 947 AA.  
 ID 09BDV3  
 AC 09BDV3  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, last annotation update)  
 DE BRCA1 (Fragment).  
 OS Dicerata; bicornis (Black rhinoceros).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Dicerata.  
 OX NCBI\_TaxID=9805;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21082081; PubMed=11214318;  
 RA Madsen O., Scally M., Douady C., Kao D., DeBry R., Adkins R.,  
 RA Amrine H., Stanhope M., de Jong W., Springer M.;  
 RT "Parallel adaptive radiations in two major clades of placental  
 RT mammals.";  
 RL Nature 409:610-614(2001).  
 DR EMBL; AF284011; AKI5598.1; -  
 DR InterPro; IPR002378; BrstCancerI.  
 DR PRINTS; PRO0493; BRSTCANCERL.  
 FT NON\_TER 1  
 FT NON\_TER 947  
 FT NON\_TER 947  
 SEQUENCE 947 AA; 104753 MW; BC8158CB0F9A404B CRC64;

Query Match 36.8%; Score 3555.5; DB 6; Length 947;  
 Best Local Similarity 75.2%; Pred. No. 1.4e-175;  
 Matches 714; Conservative 86; Mismatches 143; Indels 7; Gaps 5;  
 QY 274 CGTNTHASSLOHENSLLTKDRMNVKAEPCNKSKOPGLARSONHRMAGSKETCNDRT 333  
 Db 1 CGTNTHASSLOHENSLLTKDRMNVKAEPCNKSKOPGLARSONHRMAGSKETCNDRT 60  
 QY 334 PSTERKVDLADPLCEKEMNKOKLPCSENPRTDEVPITLNNSTIOKYNEMFSRSDLL 393  
 Db 61 PSEKVKLVNADPLGKREKLNKQKPCNSNSRSDQDVPITLNNSTIOKYNEMFSRSDLL 120

```

Db 61 PNSEKKLVNLADPLYGRRELKOKRPPCSDSPRODIDFWITRNSSIOKVNEWFSRSDTL 120
QY 394 GSDSDHDESESNKAVADVLVNEVDYSSGSEKIDLLASPHALICKSERVHSKVE 453
Db 121 TSDSDHGGPESNTEVAGAVEVNEVDYSSGSEKIDLLASPHALICKSERVHSKVE 180
QY 454 SNIEDIKFGKTYRRKASLPNLSHTVENLIIAFTVTEPOIIERPLTNLKRKRPTSGH 513
Db 181 NNIEDKIFGKTYRRKASLPNLSHTVEDLIIIGASAIESQITQERPLTNLKRKRPTSGH 240
QY 514 PEDFIKAKDLA-VOKTPEMINOGTNOTQONQVNNITSGHENKTKGSIONEKNPPIE 572
Db 241 PEDFIKAKDLA-VOKTPEMINOGTNOTQONQVNNITSGHENKTKGSIONEKNPPIE 300
QY 573 SLEKSAFRTKAEPITSSISNMNLELNHNSKAPKNNLRKRSRHHHALELVYSRNL 632
Db 301 SLEKSAFRTKAEPITSSISNMNLELNHNSKAPKNNLRKRSRHHHALELVYSRNL 360
QY 633 PPNCTELOIDSCSSSEELIKKKKYNOMPVHRNRNOLMGKBPATGAKKSNKPNQTSKRH 692
Db 361 PPNHTELOVDSCSSSEEMKKKSPQVPRHSRKLQUTEDKEPAAGAKKSNKPNQINRKL 420
QY 693 DSDFFPELKLINAGSFTKCSNTSELEFVNPSPPREKEBEKLETVKYSNNAEDPKDML 752
Db 421 ASDAFPELKLINAGSFTKCSNTSELEFVNPSPPREKEBEKLETVKYSNNAEDPKDML 480
QY 753 SGERVLOTERSVESSISLVPGTDYGTQESISLLEVSTLGAKTPEPKCVSOCAAFENPK 812
Db 481 SGGKLOAERSVESSISLVPGTDYGTQESISLLEVSTLGAKTPEPKCVSOCAAFENPK 540
QY 813 GLIHGCKDNNDNDEBGFYPLGHEVNSRSTSIEMESSELDQVLYONTFVYSKROSFALF 872
Db 541 ELIHDCSKDTRNDDEGLDPLRCEVNHQETSIEMESSELDQVLYONTFVYSKROSFALF 600
QY 873 SNPNAEDECATFSAHSGSLKKSOPKVTPECEOKENQKNSNIKPQVITAGFPV 932
Db 601 SNGSPEKCAVAHSTRSLKRPKVTLECGQKEENGKESKKAHVOSHTTVDVFPV 660
QY 933 GOKD-KPVNNAKCSIKGSRFCLSSOPFGNETGLITPNKHGLLOMPYRIPPLEPIKSEVK 991
Db 661 COKDKKPDGVKCYSTKESRLCOSQOPFGNETGLITPNKHGLLOMPYRIPPLEPIKSEVK 720
QY 992 TKCKKNLEENFEHSMSPEREMNGEN-IPSTVSTISRNINRENVFEKASSNINEVSS 1050
Db 721 TICKKNVSEKLEHSHVSPEKALGNKSVIOSTVSTISQNNIRESTFEKSSINEVSS 780
QY 1051 TNEVGSSINETSSEDENIQAELGRNRPKLNAMLRGLVLOPEVYKOSLPISNCKHPEIKR 1110
Db 781 TNEVGSSINETSSEDENIQAELGRNRPKLNAMLRGLVLOPEVYKOSLPISNCKHPEIKR 840
QY 1111 Q-EYEEVQVNTDFSPYLSIDNLEQPMGSSHASOVCEPDDLLDDEIKEDTSFAEND 1169
Db 841 RGENEGVVPYVADPSPICSDNLEQPMGSSHASOVCEPDDLLDDEIKEDTSFAEND 900
QY 1170 IKESAVFSKSVQGEUSRSPSPFTHHLAGOYRRGAKKLSESENNLSSE 1219
Db 901 IKESAVFSKSVQGEUSRSPSPFTHHLAGOYRRGAKKLSESENNLSSE 947

```

## RESULT 15

```

Q9BDV4 ID Q9BDV4 PRELIMINARY; PRT; 942 AA.
AC Q9BDV4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE BRCA1 (Fragment).
OS Equus caballus (Horse).
OC Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
OX Mammalia: Eutheria: Perissodactyla; Equidae; Equus.
RN NCBL_TaxID-9796;
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21082081; PubMed=11214318;
RA Madgen O., Scally M., Douady C., Kao D., Debry R., Adkins R.,
RA Amrine H., Stanhope M., de Jong W., Springer M.;
RT "Parallel adaptive radiations in two major clades of placental
RL mammals";
RL Nature 409:610-614(2001).
DR EMBL; AF284010; AAK15597.1;
DR InterPro; IPR002378; Brst.CancerI.
DR PRINTS; PR00493; BRSTCANCERL.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 942 AA; 104317 MW; F5C2EE1C7313C1A CRC64;

```

Query Match 36.1%; Score 3482; DB 6; Length 942;

Best Local Similarity 73.6%; Pred. No. 9e-172;

Matches 698; Conservative 90; Mismatches 151; Indels 10; Gaps 4;

```

QY 274 CGTNTAASSLOHENSLLTYKDRNVEKAEFCNKSQOPGLARSOHNWAGSKETCNDRT 333
Db 1 CGTNTAASSLOHENSLLTYKDRNVEKAEFCNKSQOPGLARSOHNWAGSKETCNDRT 60
QY 334 PSTEKKVLNADPLYCERKEKOKRPPCSDSPRODIDFWITRNSSIOKVNEWFSRSDTL 393
Db 61 PNSEKKLVNLADPLYGRRELKOKRPPCSDSPRODIDFWITRNSSIOKVNEWFSRSDTL 120
QY 394 GSDSDHDESESNKAVADVLVNEVDYSSGSEKIDLLASPHALICKSERVHSKVE 453
Db 121 TSDSDHGGPESNTEVAGAVEVNEVDYSSGSEKIDLLASPHALICKSERVHSKVE 180
QY 454 SNIEDIKFGKTYRRKASLPNLSHTVENLIIAFTVTEPOIIERPLTNLKRKRPTSGH 513
Db 181 NNIEDKIFGKTYRRKASLPNLSHTVEDLIIIGASAIESQITQERPLTNLKRKRPTSGH 240
QY 514 PEDFIKAKDLA-VOKTPEMINOGTNOTQONQVNNITSGHENKTKGSIONEKNPPIE 572
Db 241 PEDFIKAKDLA-VOKTPEMINOGTNOTQONQVNNITSGHENKTKGSIONEKNPPIE 300
QY 573 SLEKSAFRTKAEPITSSISNMNLELNHNSKAPKNNLRKRSRHHHALELVYSRNL 632
Db 301 SLEKSAFRTKAEPITSSISNMNLELNHNSKAPKNNLRKRSRHHHALELVYSRNL 360
QY 633 PPNCTELOIDSCSSSEELIKKKKYNOMPVHRNRNOLMGKBPATGAKKSNKPNQTSKRH 692
Db 361 PPNHTELOVDSCSSSEEMKKKSPQVPRHSRKLQUTEDKEPAAGAKKSNKPNQINRKL 420
QY 693 DSDFFPELKLINAGSFTKCSNTSELEFVNPSPPREKEBEKLETVKYSNNAEDPKDML 752
Db 421 ASDAFPELKLINAGSFTKCSNTSELEFVNPSPPREKEBEKLETVKYSNNAEDPKDML 480
QY 753 SGERVLOTERSVESSISLVPGTDYGTQESISLLEVSTLGAKTPEPKCVSOCAAFENPK 812
Db 481 SGGKLOAERSVESSISLVPGTDYGTQESISLLEVSTLGAKTPEPKCVSOCAAFENPK 540
QY 813 GLIHGCKDNNDNDEBGFYPLGHEVNSRSTSIEMESSELDQVLYONTFVYSKROSFALF 872
Db 541 ELIHDCSKDTRNDDEGLDPLRCEVNHQETSIEMESSELDQVLYONTFVYSKROSFALF 600
QY 873 SNPNAEDECATFSAHSGSLKKSOPKVTPECEOKENQKNSNIKPQVITAGFPV 932
Db 601 SNGSPEKCAVAHSTRSLKRPKVTLECGQKEENGKESKKAHVOSHTTVDVFPV 660
QY 933 GOKD-KPVNNAKCSIKGSRFCLSSOPFGNETGLITPNKHGLLOMPYRIPPLEPIKSEVK 991
Db 661 COKDKKPDGVKCYSTKESRLCOSQOPFGNETGLITPNKHGLLOMPYRIPPLEPIKSEVK 720
QY 992 TKCKKNLEENFEHSMSPEREMNGEN-IPSTVSTISRNINRENVFEKASSNINEVSS 1050
Db 721 TICKKNVSEKLEHSHVSPEKALGNKSVIOSTVSTISQNNIRESTFEKSSINEVSS 780
QY 1051 TNEVGSSINETSSEDENIQAELGRNRPKLNAMLRGLVLOPEVYKOSLPISNCKHPEIKR 1110
Db 775 -NEVGSSINETSSEDENIQAELGRNRPKLNAMLRGLVLOPEVYKOSLPISNCKHPEIKR 833

```

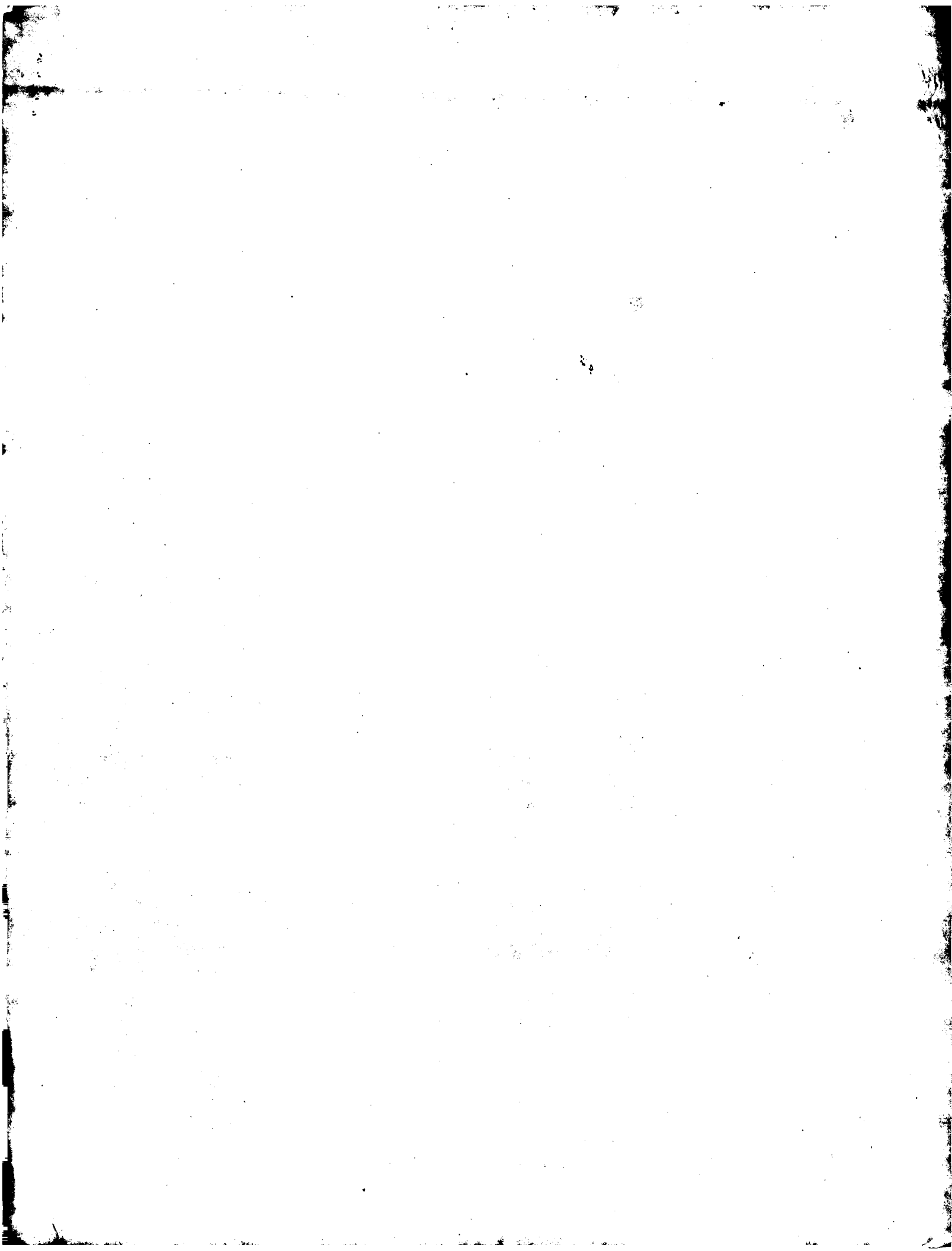
Fri Jun 27 10:47:59 2003

us-09-734-672-4.rspt

Page 15

OY	1111	QEEYEYVOTATDTPSSPLTISDNIEOMPNGSSASIVGSETPDDLLIDDEIKEDTSFAENDI	1170
		: : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Dd	834	GGEKEVVQAANADPSPCLSIDNIEQPMGSGCASGVCSCTPDLLIDDEIKENIFALSQV	893
OY	1171	KESAAVSXSKVOKGELSRSPSFHTHIAQCYRAGAKKLBSSENLSSE	1219
		: : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : :	
Dd	894	KERASVSKSVQKGRFRPSPIICRTGLCAQHQRARLRESENNTSSE	942

Search completed: June 27, 2003, 10:26:27  
Job time : 130 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 12:33:08 ; Search time 1101.77 Seconds  
(without alignments)  
11673.202 Million cell updates/sec

Title: US-09-734-672-3

Perfect score: 5711  
Sequence: 1 AGCTCGCTGACACTTCTCGG.....TCCCCACAGCCACTACTGA 5711

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%  
Maximum Match: 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:\*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	5709.4	100.0	5711 19	AAV6449
2	5709.4	100.0	5711 21	AAV6449
3	5709.4	100.0	5711 21	AAV6449
4	5709.4	100.0	5711 21	AAV6449
5	5709.4	100.0	5711 21	AAV6449
6	5709.4	100.0	5711 21	AAV6449
7	5709.4	100.0	5711 21	AAV6449
8	5709.4	100.0	5711 21	AAV6449
9	5709.4	100.0	5711 21	AAV6449

10	5709	100.0	5711 19	AAV6461
11	5709	100.0	5711 19	AAV6462
12	5707.8	99.9	5712 18	AAV6484
13	5707.8	99.9	5712 21	AAV6495
14	5707.8	99.9	5712 21	AAV6495
15	5707.8	99.9	5712 21	AAV6495
16	5707.8	99.9	5712 21	AAV6495
17	5707.8	99.9	5712 21	AAV6495
18	5706.2	99.9	5711 17	AAV64024
19	5706.2	99.9	5711 17	AAV64030
20	5706.2	99.9	5711 18	AAV6067
21	5706.2	99.9	5711 18	AAV6067
22	5706.2	99.9	5711 19	AAV60563
23	5706.2	99.9	5711 19	AAV60563
24	5706.2	99.9	5711 17	AAV6445
25	5706.2	99.9	5711 17	AAV6445
26	5706.2	99.9	5711 17	AAV6445
27	5706.2	99.9	5711 17	AAV6445
28	5706.2	99.9	5711 17	AAV6445
29	5706.2	99.9	5711 17	AAV6445
30	5706.2	99.9	5711 17	AAV6445
31	5706.2	99.9	5711 17	AAV6445
32	5706.2	99.9	5711 17	AAV6445
33	5706.2	99.9	5711 17	AAV6445
34	5706.2	99.9	5711 17	AAV6445
35	5706.2	99.9	5711 17	AAV6445
36	5706.2	99.9	5711 17	AAV6445
37	5706.2	99.9	5711 17	AAV6445
38	5706.2	99.9	5711 17	AAV6445
39	5706.2	99.9	5711 17	AAV6445
40	5706.2	99.9	5711 17	AAV6445
41	5706.2	99.9	5711 17	AAV6445
42	5706.2	99.9	5711 17	AAV6445
43	5706.2	99.9	5711 17	AAV6445
44	5706.2	99.9	5711 17	AAV6445
45	5706.2	99.9	5711 17	AAV6445

## ALIGNMENTS

RESULT 1	AAV6449	standard: cDNA, 5711 BP.
ID	AAV6449	
AC	AAV6449	
DT	18-NOV-1998	(first entry)
XX	Human BRCA1 om12 cDNA.	
XX	BRCA1, om12; human; breast and ovarian cancer predisposing gene;	
XX	polymorphism; susceptibility; anti-oncogene; tumour suppressor;	
XX	chromosome 17q; ss.	
XX	Human sapiens.	
XX	key	Location/Qualifiers
XX	CDS	120..5711
XX	FT	/tag= a
XX	FT	/product= "BRCA1 om12 protein"
XX	PN	US5750400-A.
XX	PD	12-MAY-1998.
XX	PF	12-FEB-1997;
XX	PR	12-FEB-1996;
XX	PR	12-FEB-1997;
XX	PA	(ONCO-) ONCORMED INC.

PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Schelter DB, Zeng B;  
DR MPI: 1998-296774/26.  
P-PSDB; AAM76099.

BRCA1 omi gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PS breast or ovarian cancer

PS Claim 2d: Column 41-46: 54pp: English.

This sequence encodes the human BRCA1 (breast and ovarian cancer  
CC predisposing gene) omi2 gene. This sequence and polymorphic variations of  
CC this sequence are useful for the identification of an individual who may  
CC or may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers,  
CC especially breast and ovarian cancer. It is found at human chromosome 17q  
CC which is known to be linked to cancer susceptibility, especially breast  
CC cancer. Cells containing a mutation in this gene lose the wild-type  
CC function of BRCA1 and are more susceptible to cancers.

SQ Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1383 T; 0 other;

Query Match 100.0%; Score 5709.4; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCTGCTGAGACTTCTCTGACCCCGACAGGCTGGGGTTTCTAGTAATCTGGCC 60  
DB 1 AGCTGCTGAGACTTCTCTGACCCCGACAGGCTGGGGTTTCTAGTAATCTGGCC 60  
QY 61 CCTGCGCTCAGAGGCGCTTACCTGCTGCTGGGTAAAGTTCATTGGAGACAGAA 120  
DB 61 CCTGCGCTCAGAGGCGCTTACCTGCTGCTGGGTAAAGTTCATTGGAGACAGAA 120  
QY 121 TGGATTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGCTATTATGTCAGAAA 180  
DB 121 TGGATTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGCTATTATGTCAGAAA 180  
QY 121 TGGATTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGCTATTATGTCAGAAA 180  
DB 121 TGGATTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGCTATTATGTCAGAAA 180  
QY 181 TCTTAAGAGTGTCCCATCTGCTGAGTGTGATCAAGAAAGTCTCCACAAAGTGAC 240  
DB 181 TCTTAAGAGTGTCCCATCTGCTGAGTGTGATCAAGAAAGTCTCCACAAAGTGAC 240  
QY 241 ACATATTTGGCAATTTGGCATGCTGAACCTTCTCAACGAGAAAGGCGCTTCAAGT 300  
DB 241 ACATATTTGGCAATTTGGCATGCTGAACCTTCTCAACGAGAAAGGCGCTTCAAGT 300  
QY 301 GTGCTTTATGTAAAGTATATTAACCAAAAGAGCGCTACAAAGTACAGATTAGTC 360  
DB 301 GTGCTTTATGTAAAGTATATTAACCAAAAGAGCGCTACAAAGTACAGATTAGTC 360  
QY 361 AACTTGTGAAGAGCTATTGAAATCATTTGCTTTACGCTTGACACAGTTTGGAGT 420  
DB 361 AACTTGTGAAGAGCTATTGAAATCATTTGCTTTACGCTTGACACAGTTTGGAGT 420  
QY 421 ATGCAACAGCTATATTTTGAAGAAAGAAATTAATCTCTGAAACATCTAAAGATG 480  
DB 421 ATGCAACAGCTATATTTTGAAGAAAGAAATTAATCTCTGAAACATCTAAAGATG 480  
QY 481 AAGTTTCTATCATCAAAAGTATGGGTACAGAAACCGTGCAGAAAGCTTCTACAGATG 540  
DB 481 AAGTTTCTATCATCAAAAGTATGGGTACAGAAACCGTGCAGAAAGCTTCTACAGATG 540  
QY 541 AACCCGAAATCTTCTCTGAGAGAAACAGTCTCAGTGTCAACTCTCTAACCTTGGAA 600  
DB 541 AACCCGAAATCTTCTCTGAGAGAAACAGTCTCAGTGTCAACTCTCTAACCTTGGAA 600  
QY 601 CTGTGGAAGTCTGAGAGCAAGGAGGATTAACCTCAAAAGAGCTGTCTACATTG 660  
DB 601 CTGTGGAAGTCTGAGAGCAAGGAGGATTAACCTCAAAAGAGCTGTCTACATTG 660

QY 661 AATTGGAGTCTGATTTCTTGAAGATACGTTTAAATAGCAACTTATTCAGTGTGGAG 720  
DB 661 AATTGGAGTCTGATTTCTTGAAGATACGTTTAAATAGCAACTTATTCAGTGTGGAG 720  
QY 721 ATCAGAAATTTGTAACAATCACCCTCAAGAGAACAGGATGAATCACTTTGATCTG 780  
DB 721 ATCAGAAATTTGTAACAATCACCCTCAAGAGAACAGGATGAATCACTTTGATCTG 780  
QY 781 CAAAAAGGCTGCTGTAATTTTGAAGAGGATGATCAAAATCTGAACATCATCAAC 840  
DB 781 CAAAAAGGCTGCTGTAATTTTGAAGAGGATGATCAAAATCTGAACATCATCAAC 840  
QY 841 CCAATATATGATTTGAACACACAGAGGCTGACAGGATGAGAGGATTCAGAAAGT 900  
DB 841 CCAATATATGATTTGAACACACAGAGGCTGACAGGATGAGAGGATTCAGAAAGT 900  
QY 901 ATCAGGATGCTGCTGTTCAACTTCATGATGAGGATGAGGATGAGGATGAGGAT 960  
DB 901 ATCAGGATGCTGCTGTTCAACTTCATGATGAGGATGAGGATGAGGATGAGGAT 960  
QY 961 GCTCATATGAGATGAGAGAGAGGATTTATCTACTAAGACAGAAATGATGAGAA 1020  
DB 961 GCTCATATGAGATGAGAGAGAGGATTTATCTACTAAGACAGAAATGATGAGAA 1020  
QY 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCTGCTTAAAGAGAGGATCAATACAGAT 1080  
DB 1021 AGGCTGAATTTCTGTAATTAAGCAACAGCTGCTTAAAGAGAGGATCAATACAGAT 1080  
QY 1081 GGGCTGGAATTAAGGAACATGTAATGATAGGAGGATCCAGACAGAAAAAGGATG 1140  
DB 1081 GGGCTGGAATTAAGGAACATGTAATGATAGGAGGATCCAGACAGAAAAAGGATG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGAGAGAAAGATGATGAGGATGAGGATGAGGAT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGAGAGAAAGATGATGAGGATGAGGATGAGGAT 1200  
QY 1201 CAGAGAAATCTGAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1260  
DB 1201 CAGAGAAATCTGAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1260  
QY 1261 AAGTTAATGAGGATTTTCCAGAGATGATGAGGATGAGGATGAGGATGAGGAT 1320  
DB 1261 AAGTTAATGAGGATTTTCCAGAGATGATGAGGATGAGGATGAGGATGAGGAT 1320  
QY 1321 GGGAGTCTGATCAATATGCAAGATGATGATGATGATGATGATGATGATGATGAT 1380  
DB 1321 GGGAGTCTGATCAATATGCAAGATGATGATGATGATGATGATGATGATGATGAT 1380  
QY 1381 AATATTTCTGCTTCTGAGAGAAATAGACTTACTGAGCAGTATGATGAGGCTTAA 1440  
DB 1381 AATATTTCTGCTTCTGAGAGAAATAGACTTACTGAGCAGTATGATGAGGCTTAA 1440  
QY 1441 TATGTAAAGTGAAGAGTCTCAATCAATCAGTATGATGATGATGATGATGAT 1500  
DB 1441 TATGTAAAGTGAAGAGTCTCAATCAATCAGTATGATGATGATGATGATGAT 1500  
QY 1501 TTGGGAAAGCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
DB 1501 TTGGGAAAGCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
QY 1561 TAAATTTAGAGCATTTGTTACTGAGCAGATTAATACAGAGGCTCCCTCACAATA 1620  
DB 1561 TAAATTTAGAGCATTTGTTACTGAGCAGATTAATACAGAGGCTCCCTCACAATA 1620  
QY 1621 AATTAAAGGCTTAAGAGAGAGTCAATCAATCAGTATGATGATGATGATGATGAT 1680  
DB 1621 AATTAAAGGCTTAAGAGAGAGTCAATCAATCAGTATGATGATGATGATGATGAT 1680  
QY 1681 CAGATTTGAGAGTTCAGAAAGAGTCTGAAATGATTAATCAGGAGTCAACCAAGGAG 1740  
DB 1681 CAGATTTGAGAGTTCAGAAAGAGTCTGAAATGATTAATCAGGAGTCAACCAAGGAG 1740

QY 1741 AGAATGTCAGATGAAATATTACTAATAGTGTGATGAGAAATAAACAAGGTGATT 1800  
 1741 AGAATGTCAGATGAAATATTACTAATAGTGTGATGAGAAATAAACAAGGTGATT 1800  
 QY 1801 CTATTCAGAAATGAGAAATCTTAAACCAATAGATACCTCGAAGAAAGATCTGCTTCA 1860  
 1801 CTATTCAGAAATGAGAAATCTTAAACCAATAGATACCTCGAAGAAAGATCTGCTTCA 1860  
 Db 1801 CTATTCAGAAATGAGAAATCTTAAACCAATAGATACCTCGAAGAAAGATCTGCTTCA 1860  
 QY 1861 AAACGAAGCTGACCTATTAAGAGAGAGTAAAGCAATATGAACTGAAATTAATATTC 1920  
 1861 AAACGAAGCTGACCTATTAAGAGAGAGTAAAGCAATATGAACTGAAATTAATATTC 1920  
 Db 1861 AAACGAAGCTGACCTATTAAGAGAGAGTAAAGCAATATGAACTGAAATTAATATTC 1920  
 QY 1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAGTCTTACAGGCAATATTC 1980  
 1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAGTCTTACAGGCAATATTC 1980  
 Db 1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAGTCTTACAGGCAATATTC 1980  
 QY 1981 ATGCGCTGAACTAGTATGATGAGAAATTAAGCCCACTAATTTGACTGAATTCGAA 2040  
 1981 ATGCGCTGAACTAGTATGATGAGAAATTAAGCCCACTAATTTGACTGAATTCGAA 2040  
 Db 1981 ATGCGCTGAACTAGTATGATGAGAAATTAAGCCCACTAATTTGACTGAATTCGAA 2040  
 QY 2041 TTGATAGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACAAACCAATGCCAGTCA 2100  
 2041 TTGATAGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACAAACCAATGCCAGTCA 2100  
 Db 2041 TTGATAGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACAAACCAATGCCAGTCA 2100  
 QY 2101 GGCACAGCAGAAACCTACACATCATGGAAGTAAAGAACTGCACACTGAGCCAGAGAGA 2160  
 2101 GGCACAGCAGAAACCTACACATCATGGAAGTAAAGAACTGCACACTGAGCCAGAGAGA 2160  
 Db 2101 GGCACAGCAGAAACCTACACATCATGGAAGTAAAGAACTGCACACTGAGCCAGAGAGA 2160  
 QY 2161 GTACACAGCAATTAAGAAAGAGAGAGTAAAGAGATGACAGTACTTCCAGAGCTGA 2220  
 2161 GTACACAGCAATTAAGAAAGAGAGTAAAGAGATGACAGTACTTCCAGAGCTGA 2220  
 Db 2161 GTACACAGCAATTAAGAAAGAGAGTAAAGAGATGACAGTACTTCCAGAGCTGA 2220  
 QY 2221 AGTTACAAATGACAGCTGTTCTTTACTAAGTGTCAATACCAGTGAATTAAGAAAT 2280  
 2221 AGTTACAAATGACAGCTGTTCTTTACTAAGTGTCAATACCAGTGAATTAAGAAAT 2280  
 Db 2221 AGTTACAAATGACAGCTGTTCTTTACTAAGTGTCAATACCAGTGAATTAAGAAAT 2280  
 QY 2281 TTGTCAATCTAGCTTCCAGAGAGAAAGAAAGAGAAAGTAAAGCAAGTTAAAGTGT 2340  
 2281 TTGTCAATCTAGCTTCCAGAGAGAAAGAAAGAGAAAGTAAAGCAAGTTAAAGTGT 2340  
 Db 2281 TTGTCAATCTAGCTTCCAGAGAGAAAGAAAGAGAAAGTAAAGCAAGTTAAAGTGT 2340  
 QY 2341 CTATTAATGCTGAAG 2400  
 2341 CTATTAATGCTGAAG 2400  
 Db 2341 CTATTAATGCTGAAG 2400  
 QY 2401 AAAGATCTGAG 2460  
 2401 AAAGATCTGAG 2460  
 Db 2401 AAAGATCTGAG 2460  
 QY 2461 AAAGATCTGAG 2520  
 2461 AAAGATCTGAG 2520  
 Db 2461 AAAGATCTGAG 2520  
 QY 2521 GGTGAGTGTGAG 2580  
 2521 GGTGAGTGTGAG 2580  
 Db 2521 GGTGAGTGTGAG 2580  
 QY 2581 ATAAATGAATGACAG 2640  
 2581 ATAAATGAATGACAG 2640  
 Db 2581 ATAAATGAATGACAG 2640  
 QY 2641 GGGAAACAAAGCATAGAAATGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2700  
 2641 GGGAAACAAAGCATAGAAATGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2700  
 Db 2641 GGGAAACAAAGCATAGAAATGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2700  
 QY 2701 TCAAGGTTTCAAGGAG 2760  
 2701 TCAAGGTTTCAAGGAG 2760  
 Db 2701 TCAAGGTTTCAAGGAG 2760  
 QY 2761 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2820  
 2761 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2820  
 Db 2761 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2820  
 QY 2820 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2880  
 2820 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2880  
 Db 2820 AATGTCAGCAATTTCTGCGCACTGCGGCTTAAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2880

Db 2821 TTGATGTCAGAAAG 2880  
 QY 2881 AGACAGTTAATATTCAGTGCAGAGGCTTCTGCTGTTGTCAGAAAGATTAAGCAAGTATA 2940  
 2881 AGACAGTTAATATTCAGTGCAGAGGCTTCTGCTGTTGTCAGAAAGATTAAGCAAGTATA 2940  
 Db 2881 AGACAGTTAATATTCAGTGCAGAGGCTTCTGCTGTTGTCAGAAAGATTAAGCAAGTATA 2940  
 QY 2941 ATGCAAAATGATATTAAG 3000  
 2941 ATGCAAAATGATATTAAG 3000  
 Db 2941 ATGCAAAATGATATTAAG 3000  
 QY 3001 ACAGAACTGAGCTATTAATTAAG 3060  
 3001 ACAGAACTGAGCTATTAATTAAG 3060  
 Db 3001 ACAGAACTGAGCTATTAATTAAG 3060  
 QY 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATTAAGTAAAGAAATCTGCTAGAGG 3120  
 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATTAAGTAAAGAAATCTGCTAGAGG 3120  
 Db 3061 CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATTAAGTAAAGAAATCTGCTAGAGG 3120  
 QY 3121 AAAACTTTGAGAAACATTAATGTCACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
 3121 AAAACTTTGAGAAACATTAATGTCACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
 Db 3121 AAAACTTTGAGAAACATTAATGTCACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
 QY 3181 GTACAGTGAACAGATTAAG 3240  
 3181 GTACAGTGAACAGATTAAG 3240  
 Db 3181 GTACAGTGAACAGATTAAG 3240  
 QY 3241 CAAGCAATTAATTAAG 3300  
 3241 CAAGCAATTAATTAAG 3300  
 Db 3241 CAAGCAATTAATTAAG 3300  
 QY 3301 TAGGTTCCAGTGTGAAAG 3360  
 3301 TAGGTTCCAGTGTGAAAG 3360  
 Db 3301 TAGGTTCCAGTGTGAAAG 3360  
 QY 3361 ATGCTATGCTTAATTAAG 3420  
 3361 ATGCTATGCTTAATTAAG 3420  
 Db 3361 ATGCTATGCTTAATTAAG 3420  
 QY 3421 GTATTTGTAAGCATCTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
 3421 GTATTTGTAAGCATCTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
 Db 3421 GTATTTGTAAGCATCTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
 QY 3481 ATACAGATTTCTCTCCATATCTGATTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 3481 ATACAGATTTCTCTCCATATCTGATTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 Db 3481 ATACAGATTTCTCTCCATATCTGATTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
 QY 3541 ATGCAATCTCAGGTTTGTGAG 3600  
 3541 ATGCAATCTCAGGTTTGTGAG 3600  
 Db 3541 ATGCAATCTCAGGTTTGTGAG 3600  
 QY 3601 AAGATCTAGTTTGTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
 3601 AAGATCTAGTTTGTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
 Db 3601 AAGATCTAGTTTGTGGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
 QY 3661 TCCAG 3720  
 3661 TCCAG 3720  
 Db 3661 TCCAG 3720  
 QY 3721 GTTACCGAAG 3780  
 3721 GTTACCGAAG 3780  
 Db 3721 GTTACCGAAG 3780  
 QY 3781 AAGAGCTTCCCTGCTTCAACAGCTTTAATTTGTAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3840  
 3781 AAGAGCTTCCCTGCTTCAACAGCTTTAATTTGTAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3840  
 Db 3781 AAGAGCTTCCCTGCTTCAACAGCTTTAATTTGTAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3840  
 QY 3841 CTACTAGGAGATGACAG 3900  
 3841 CTACTAGGAGATGACAG 3900  
 Db 3841 CTACTAGGAGATGACAG 3900  
 QY 3901 TATCATTTGAGAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3960  
 3901 TATCATTTGAGAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3960  
 Db 3901 TATCATTTGAGAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3960

Dp	3901	TATCATTTGAAGAAATAGCTTAAATGACTGCAGTAACCAAGTAAATATTTGGCAAAAGCATCTC	3960
Qy	3961	AGGAACATTCACCTTATGTCAGAGAACAAATGTTCTGTACGCTGTTTCTTCACAGTSCA	4020
Dp	3961	AGGAACATTCACCTTATGTCAGAGAACAAATGTTCTGTACGCTGTTTCTTCACAGTSCA	4020
Qy	4021	GTGAATTTGGAAACACTTGAAGTCAAAATACAAACACCAGATCCCTTCTTGATTTGGTCTT	4080
Dp	4021	GTGAATTTGGAAACACTTGAAGTCAAAATACAAACACCAGATCCCTTCTTGATTTGGTCTT	4080
Qy	4081	CCAAACAAATGAGGACATGCTCGAAAGCCAGGAGATTGGTCTGAGTACAGGAATTTGG	4140
Dp	4081	CCAAACAAATGAGGACATGCTCGAAAGCCAGGAGATTGGTCTGAGTACAGGAATTTGG	4140
Qy	4141	TTTCAGATGTATGAACAAAGAGAGAACGGGCTTTGGAAAGAAATATCAACAAAGCAACCA	4200
Dp	4141	TTTCAGATGTATGAACAAAGAGAGAACGGGCTTTGGAAAGAAATATCAACAAAGCAACCA	4200
Qy	4201	TGGAATTCAAACTTATGATGAGAGCAGCATCTGGGTGTGAGAGTGAACCAAGCCTCTCTGAG	4260
Dp	4201	TGGAATTCAAACTTATGATGAGAGCAGCATCTGGGTGTGAGAGTGAACCAAGCCTCTCTGAG	4260
Qy	4261	ACTGCTCAGGGCTATTCCTCTCAGAGTCAACATTTTAAACCCTCAGCAGAGGATACCATGC	4320
Dp	4261	ACTGCTCAGGGCTATTCCTCTCAGAGTCAACATTTTAAACCCTCAGCAGAGGATACCATGC	4320
Qy	4321	AACATTAACCTGATTAAGCTCCAGCAGGAAATGGCGAATCGAATCGAAGCCTGTAGAACAGC	4380
Dp	4321	AACATTAACCTGATTAAGCTCCAGCAGGAAATGGCGAATCGAATCGAAGCCTGTAGAACAGC	4380
Qy	4381	ATGGAGCCAGGCTTCTTAACAGCTACCCCTCCATCATATAGTACTCTTCTGCCCTTGAGG	4440
Dp	4381	ATGGAGCCAGGCTTCTTAACAGCTACCCCTCCATCATATAGTACTCTTCTGCCCTTGAGG	4440
Qy	4441	ACCTCGCAATTCACAAACAAAGACATCAGAAAAAGCGATPTTAACCTTCACAGAAAAAGTA	4500
Dp	4441	ACCTCGCAATTCACAAACAAAGACATCAGAAAAAGCGATPTTAACCTTCACAGAAAAAGTA	4500
Qy	4501	GTGAATACCCCTATAAGCCAGAAATCCAGAAAGGCTTTCGCTGACAAAGTTTGAGGTGCTG	4560
Dp	4501	GTGAATACCCCTATAAGCCAGAAATCCAGAAAGGCTTTCGCTGACAAAGTTTGAGGTGCTG	4560
Qy	4561	CAGATAGTTCATCAACGATTAATAAATAAGAACCAAGAGTGTGAAAGATCATCCCCCTCTAAT	4620
Dp	4561	CAGATAGTTCATCAACGATTAATAAATAAGAACCAAGAGTGTGAAAGATCATCCCCCTCTAAT	4620
Qy	4621	GCCCATCATTTAGATAGTATAGTGTGCATGACAGTGTCTGCGGAGTCTTCAGATATGAA	4680
Dp	4621	GCCCATCATTTAGATAGTATAGTGTGCATGACAGTGTCTGCGGAGTCTTCAGATATGAA	4680
Qy	4681	ACTACCCATCTCAGAGAGAGCTCATTAAGGTTGTTGATGTGAGAGAGCACAGCTCGAAG	4740
Dp	4681	ACTACCCATCTCAGAGAGAGCTCATTAAGGTTGTTGATGTGAGAGAGCACAGCTCGAAG	4740
Qy	4741	AGTCGGGACACACATTTGACGGAACATCTTACTTCCCAAGGCAAGATCTTAAGGGAA	4800
Dp	4741	AGTCGGGACACACATTTGACGGAACATCTTACTTCCCAAGGCAAGATCTTAAGGGAA	4800
Qy	4801	CCCCCTTACCTGGAATCTGGAATTCAGCCCTTCTCTGATGACCCCTGAATCTGATCCTTCTG	4860
Dp	4801	CCCCCTTACCTGGAATCTGGAATTCAGCCCTTCTCTGATGACCCCTGAATCTGATCCTTCTG	4860
Qy	4861	AAGACAGAGCCCCAGAGTCAGCTGTGTTGGCAACATACCATCTTCAACCTCTGCATTTGA	4920
Dp	4861	AAGACAGAGCCCCAGAGTCAGCTGTGTTGGCAACATACCATCTTCAACCTCTGCATTTGA	4920
Qy	4921	AAAGTTCGCCCAATTTGAAGTTGCAAGATCTGGCCACAGAGTCACAGTCGCTCATACTCTG	4980
Dp	4921	AAAGTTCGCCCAATTTGAAGTTGCAAGATCTGGCCACAGAGTCACAGTCGCTCATACTCTG	4980
Qy	4981	ATACTGCTGGGTATATATCAATTCGAGAAAGTGTGACGAGGAGAGACCGAATTTGACAG	5040
Dp	4981	ATACTGCTGGGTATATATCAATTCGAGAAAGTGTGACGAGGAGAGACCGAATTTGACAG	5040

QY	5041	CTTCAACAGAAAGGCTCAACAAAGAATGTCATCGTGTGTCGGCTGACCCCAAG	5100
Db	5041	CTTCAACAGAAAGGCTCAACAAAGAAATGTCATCGTGTGTCGGCTGACCCCAAG	5100
QY	5101	AATTATCTCGTGTACAAAGTTTGGCAGAAAAACACACATCATCTTAATTAATTA	5160
Db	5101	AATTATCTCGTGTACAAAGTTTGGCAGAAAAACACACATCATCTTAATTAATTA	5160
QY	5161	CTGAAGAGACTACTCATCTGTTATGAAGACAGATGCTGAGTTGTGTGAACGACAC	5220
Db	5161	CTGAAGAGACTACTCATCTGTTATGAAGACAGATGCTGAGTTGTGTGAACGACAC	5220
QY	5221	TGAATATATTTCTAGAGATTGCGGGAGAGAAATGGTAGTATTTCTGGGTGACC	5280
Db	5221	TGAATATATTTCTAGAGATTGCGGGAGAGAAATGGTAGTATTTCTGGGTGACC	5280
QY	5281	AGTCATTTAAAGAAAGAAATGCTGATGAGCATGATTTGAAGTCAGAGAGATGTGG	5340
Db	5281	AGTCATTTAAAGAAAGAAATGCTGATGAGCATGATTTGAAGTCAGAGAGATGTGG	5340
QY	5341	TCAATGAGAAAGAACCCACCAAGGTCCTCAAGCGAGAGAGAAATCCAGAGACGAAAGATCT	5400
Db	5341	TCAATGAGAAAGAACCCACCAAGGTCCTCAAGCGAGAGAGAAATCCAGAGACGAAAGATCT	5400
QY	5401	TCAGGGGGCTGAAATCTGTTGCTATGAGGGCCCTTACCAACATGCCACAGATCACTGG	5460
Db	5401	TCAGGGGGCTGAAATCTGTTGCTATGAGGGCCCTTACCAACATGCCACAGATCACTGG	5460
QY	5461	AATGATGTGTACAGCTGTGTGGTCTTCTGTGTGAAGAGACCTTTCATTCATCCCTTG	5520
Db	5461	AATGATGTGTACAGCTGTGTGGTCTTCTGTGTGAAGAGACCTTTCATTCATCCCTTG	5520
QY	5521	GCACAGGTGTCCACCAATTTGTGTGTGACGCCAGATGCTGTGACAGAGACAAATGGCT	5580
Db	5521	GCACAGGTGTCCACCAATTTGTGTGTGACGCCAGATGCTGTGACAGAGACAAATGGCT	5580
QY	5581	TCATAGCAATTTGGGACAGATGTGTGAGGACACTGTGTGAGACCCGAGAGAGGGGTGTGGACA	5640
Db	5581	TCATAGCAATTTGGGACAGATGTGTGAGGACACTGTGTGAGACCCGAGAGAGGGGTGTGGACA	5640
QY	5641	GTTGAGCACTGTACACAGTGCAGAGAGCTGACACCTACCTGATACCCAGATGCCGTGTTGGACA	5700
Db	5641	GTTGAGCACTGTACACAGTGCAGAGAGCTGACACCTACCTGATACCCAGATGCCGTGTTGGACA	5700
QY	5701	GCCACTACTGA 5711	
Db	5701	GCCACTACTGA 5711	
RESULT 2			
AAC60795			
ID	AAC60795 standard; cDNA: 5711 BP.		
XX			
AC	AAC60795;		
XX			
DT	07-FEB-2001 (first entry)		
XX			
DE	Human BRCA1 (om13) nucleotide sequence SEQ ID NO:5.		
XX			
KW	Human; BRCA1; chromosome 17; 17q21; breast cancer; ovarian cancer;		
KW	gene therapy; diagnosis; cytostatic; genetic susceptibility; mutation;		
KW	polymorphism; identification; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	US6130322-A.		
XX			
PD	10-OCT-2000.		
XX			
PF	06-MAY-1998; 98US-0074476.		
XX			
PR	12-FEB-1996; 96US-0598591.		



PR 12-DEC-1997; 9705-0798691.

XX (GENE-) GENE LOGIC INC.

XX Zeng B, Thuber D, Olson SJ, Alvares CP, Allen ACP, Murphy PD;  
PI Critz BS;

XX WPI: 2000-646756/62.  
XX P-PSDB: AAB24219.

DR New coding sequence of the human BRCA1 gene, i.e. BRCA1 (om12), useful  
PT in gene therapy, especially for preventing or treating breast or  
PT ovarian cancer, as well as for diagnosing or monitoring breast or  
PT ovarian cancer

Example 4: Column 59-66; 56pp; English.

CC AAC60793 to AAC60795 encode the human BRCA1 (om1-3) proteins given in  
CC AAB24217 to AAB24219 respectively. BRCA1 is found on chromosome 17  
CC mapping to position 17q21. The BRCA1 (om12) coding sequence is  
CC specifically claimed in the present invention. The BRCA1 (om12) coding  
CC sequence is useful in gene therapy, especially for preventing or treating  
CC breast or ovarian cancer. It is also useful for diagnosing or monitoring  
CC breast or ovarian cancer. Furthermore, the BRCA1 (om12) coding sequence  
CC is useful for: (a) identifying individuals having BRCA1 gene mutations  
CC and having an increased genetic susceptibility to breast or ovarian  
CC cancer, or identifying a mutation that increases the genetic  
CC susceptibility to breast or ovarian cancer; (b) avoiding  
CC misinterpretation of polymorphisms found in the BRCA1 gene; (c)  
CC determining the presence of a previously unknown mutation in the BRCA1  
CC gene; (d) probing a human sample of the BRCA1 gene by allele to determine  
CC the presence of either polymorphic alleles or mutations; and (e)  
CC performing diagnosis with a reagent derived from the BRCA1 (om1) cDNA  
CC sequence. AAC60796 to AAC60861 represent PCR primers for the BRCA1 gene,  
CC which are used in an example from the present invention.

XX Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1383 T; 0 other;

XX Query Match 100.0%; Score 5709.4; DB 21; Length 5711;

XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCTGCTGAGACTTCTGTGACCCGACAGGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
DB 1 AGCTGCTGAGACTTCTGTGACCCGACAGGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
QY 61 CCTGCGCTCAGAGAGGCTTCCACCTGCTGGGTAAGTCAATTGGAAGAAAGAA 120  
DB 61 CCTGCGCTCAGAGAGGCTTCCACCTGCTGGGTAAGTCAATTGGAAGAAAGAA 120  
QY 121 TGAATTTATCTGCTTCTGCGCTTGAAGAGTCAAAATGTCATTAATGTCAGAGAAA 180  
DB 121 TGAATTTATCTGCTTCTGCGCTTGAAGAGTCAAAATGTCATTAATGTCAGAGAAA 180  
QY 181 TCTTAGAGTGTCCATCTGTGTGAGTTGATCAAGAAAGCTGTCTCCAAAGTGTGACC 240  
DB 181 TCTTAGAGTGTCCATCTGTGTGAGTTGATCAAGAAAGCTGTCTCCAAAGTGTGACC 240  
QY 241 ACATATTTTGGCAATTTTGCATGCTGAAACTTCAACCAAGAAAGAGGCTTCCACAGT 300  
DB 241 ACATATTTTGGCAATTTTGCATGCTGAAACTTCAACCAAGAAAGAGGCTTCCACAGT 300  
QY 301 GTCTTTATGTAGAAATGATTAACCAAAAGAGGCTTCAAGAAAGTACGAGATTGATGTC 360  
DB 301 GTCTTTATGTAGAAATGATTAACCAAAAGAGGCTTCAAGAAAGTACGAGATTGATGTC 360  
QY 361 AACTTGTGAAGAGCTATGAAATCATTTGTCTTTTACGCTTGACACAGGTTTGAGT 420  
DB 361 AACTTGTGAAGAGCTATGAAATCATTTGTCTTTTACGCTTGACACAGGTTTGAGT 420  
QY 421 ATGCAAAAGCTATTAATTTTGCAGAAAAGAAATTAATCTCCGGAACATTAAGATG 480  
DB 421 ATGCAAAAGCTATTAATTTTGCAGAAAAGAAATTAATCTCCGGAACATTAAGATG 480

QY 481 AAGTTTCTATCATCCAAAGTATGGGTACAGAAACCGTCCAAAGACTTCTACAGATG 540  
DB 481 AAGTTTCTATCATCCAAAGTATGGGTACAGAAACCGTCCAAAGACTTCTACAGATG 540  
QY 541 AACCAGAAATCTTCTGTCAGAGAAACCAAGTCTCAGTCTCCACTCTTAACTTTGAA 600  
DB 541 AACCAGAAATCTTCTGTCAGAGAAACCAAGTCTCAGTCTCCACTCTTAACTTTGAA 600  
QY 601 CTGTGAGAACTCTGAGGACAAAGCAGGATCAACCTCAAAAGAGCTGTCTACATTG 660  
DB 601 CTGTGAGAACTCTGAGGACAAAGCAGGATCAACCTCAAAAGAGCTGTCTACATTG 660  
QY 661 AATTGGATCTGATCTTCTGAGAGTACCGTTAATTAAGCACTTAATGCAATGGGAG 720  
DB 661 AATTGGATCTGATCTTCTGAGAGTACCGTTAATTAAGCACTTAATGCAATGGGAG 720  
QY 721 ATCAAGAAATTTTCAAAATCAACCCCTCAAGAAACGAGGATGAATTCAGTTTGTG 780  
DB 721 ATCAAGAAATTTTCAAAATCAACCCCTCAAGAAACGAGGATGAATTCAGTTTGTG 780  
QY 781 CAAAAAGGCTGCTTGTGATTTTCTGAGAGGATGAACAAATCTGATCATCAAC 840  
DB 781 CAAAAAGGCTGCTTGTGATTTTCTGAGAGGATGAACAAATCTGATCATCAAC 840  
QY 841 CCAGTAATTAATGATTTGAACCCACTGAGAGGCGTGTGAGGCGATCCAGAAAGT 900  
DB 841 CCAGTAATTAATGATTTGAACCCACTGAGAGGCGTGTGAGGCGATCCAGAAAGT 900  
QY 901 ATCAGAGTATGTTCTGTTTCAAACTTGCATGTGAGGAGCCATGTCATCTATGCA 960  
DB 901 ATCAGAGTATGTTCTGTTTCAAACTTGCATGTGAGGAGCCATGTCATCTATGCA 960  
QY 961 GCTCATTCACAGATGAGAAACAGCACTTTATTAATCTAATTAAGCAAGATTAAGTGA 1020  
DB 961 GCTCATTCACAGATGAGAAACAGCACTTTATTAATCTAATTAAGCAAGATTAAGTGA 1020  
QY 1021 AAGCTGAATTTCTGATTAATTAAGCAAGGCGGCTTGAAGAGGCGCAACATCAAGT 1080  
DB 1021 AAGCTGAATTTCTGATTAATTAAGCAAGGCGGCTTGAAGAGGCGCAACATCAAGT 1080  
QY 1081 GGGCTGGAAGTAAAGAAACATGTAATGATGAGCGGAGTCCAGCAGCAAAAAAGGTAG 1140  
DB 1081 GGGCTGGAAGTAAAGAAACATGTAATGATGAGCGGAGTCCAGCAGCAAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGAAGCAAGAACTGCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGAAGCAAGAACTGCATGCT 1200  
QY 1201 CAGAGAAATCTAGAGATCTGAAAGATGTTCTTGAATTAAGCAAGCAATTTGAGA 1260  
DB 1201 CAGAGAAATCTAGAGATCTGAAAGATGTTCTTGAATTAAGCAAGCAATTTGAGA 1260  
QY 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTGAATGAACTGTGAGTCACTACATGATG 1320  
DB 1261 AAGTTAATGAGTGTGTTTCCAGAAAGTGAATGAACTGTGAGTCACTACATGATG 1320  
QY 1321 GGGAGTCTGAATCAAAATGCCAAAGTGAATGATTAATGAGGAGTGAAGT 1380  
DB 1321 GGGAGTCTGAATCAAAATGCCAAAGTGAATGATTAATGAGGAGTGAAGT 1380  
QY 1381 AATATTTCTGTTCTTCAAGAGAAATAGACTTACAGGAGTCACTATGAGGCTTTAA 1440  
DB 1381 AATATTTCTGTTCTTCAAGAGAAATAGACTTACAGGAGTCACTATGAGGCTTTAA 1440  
QY 1441 TATGTAAGAGTGAAGAGTCTCAGTCAATCAGTGAAGAGTGAATTAATTAAGCAAAATAT 1500  
DB 1441 TATGTAAGAGTGAAGAGTCTCAGTCAATCAGTGAAGAGTGAATTAATTAAGCAAAATAT 1500  
QY 1501 TTGGGAAAGAACTATTCGGAAGAGCAAGCCCTCCCAACTTAAGCATTAAGCAATTC 1560  
DB 1501 TTGGGAAAGAACTATTCGGAAGAGCAAGCCCTCCCAACTTAAGCATTAAGCAATTC 1560

OY	1561	TAAATTATGGAGCACTTTGTTCTCTGCGCCACAGATTAATCAAGAGCGTCCCTCCACAATA	162
Db	1561	TAATTATGGACATTTGTTACTGTGCGCCACGATATATCAAGAGCGTCCCTCCACAATA	1620
OY	1621	TAATTAAACGGTAAAGAGACCTACATAGGCGCTTCATCTGAGATTTTATCAAGAAAG	1680
Db	1621	AATTAAACCGTAAAGAGACCTACATCAGGCGCTTCATCTGAGATTTTATCAAGAAAG	1680
OY	1661	CAGATTTGGCACTTAAAAAGACTCCTGAAATGATTAATCAGGAACTAACCAAGGAGC	1740
Db	1661	CAGATTTGGCACTTAAAAAGACTCCTGAAATGATTAATCAGGAACTAACCAAGGAGC	1740
OY	1741	AGAAATGCTCAAGTGCATGAATATTACTAATAGTGGTCAATGAGAAATAAACAAGAGGTAT	1800
Db	1741	AGAAATGCTCAAGTGCATGAATATTACTAATAGTGGTCAATGAGAAATAAACAAGAGGTAT	1800
OY	1801	CTATTTCAGAAATGAGAAAAATCTTAACCAATATGATCTCTGAAAAAGAAATCTGTTTCA	1860
Db	1801	CTATTTCAGAAATGAGAAAAATCTTAACCAATATGATCTCTGAAAAAGAAATCTGTTTCA	1860
OY	1861	AAACGAAGCTGGAACCTTAAGACAGCAGTATTAAGCAATATGGAATCGAATTAATATCC	1920
Db	1861	AAACGAAGCTGGAACCTTAAGACAGCAGTATTAAGCAATATGGAATCGAATTAATATCC	1920
OY	1921	ACAATTCAAAAGCACCTTAAAAAGATAGGCTGAGAGAGAAGTCTTTACCAAGCATATTC	1980
Db	1921	ACAATTCAAAAGCACCTTAAAAAGATAGGCTGAGAGAGAAGTCTTTACCAAGCATATTC	1980
OY	1981	ATGCGCTTGAAGTACTAGTACAGTACAAATCTAAGCCCACTAATATGTACTCAATTTGCANA	2040
Db	1981	ATGCGCTTGAAGTACTAGTACAGTACAAATCTAAGCCCACTAATATGTACTCAATTTGCANA	2040
OY	2041	TTGATAGTTGTTCTAGACAGTGAAGATTAAGAAAAAAAAGTACACCAATAGCCAGTCA	2100
Db	2041	TTGATAGTTGTTCTAGACAGTGAAGATTAAGAAAAAAAAGTACACCAATAGCCAGTCA	2100
OY	2101	GGCAGAGAGAAACCTTACAACCTCAGTGAAGGTAAGAAAGCTGCAACTGGAGCCAAAGAGA	2160
Db	2101	GGCAGAGAGAAACCTTACAACCTCAGTGAAGGTAAGAAAGCTGCAACTGGAGCCAAAGAGA	2160
OY	2161	GTAACAAACCCAAATGACAGACAGATTAAGAAACATGACAGTATCTTTCCAGAGCTGA	2220
Db	2161	GTAACAAACCCAAATGACAGACAGATTAAGAAACATGACAGTATCTTTCCAGAGCTGA	2220
OY	2221	AGTTTACAAATGACACCTGCTGTTCTTTACTAAGTGTTCAAATACAGTAACTTAAGAT	2280
Db	2221	AGTTTACAAATGACACCTGCTGTTCTTTACTAAGTGTTCAAATACAGTAACTTAAGAT	2280
OY	2281	TTGTCAATCTAGACCTTCCAAAGAGAAAGAAAGAAACATGAAACAGTTTAAAGGT	2340
Db	2281	TTGTCAATCTAGACCTTCCAAAGAGAAAGAAAGAAACATGAAACAGTTTAAAGGT	2340
OY	2341	CTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTGCAAACTG	2400
Db	2341	CTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTGCAAACTG	2400
OY	2401	AAAAGTCTGTAGAGGTAGCAGATTTTCAATGTAACCTGATACGATTAATGCGCACTCAG	2460
Db	2401	AAAAGTCTGTAGAGGTAGCAGATTTTCAATGTAACCTGATACGATTAATGCGCACTCAG	2460
OY	2461	AAAGTATCTGTTACTGGAAGTTAGCACCTTAGGGAAGGCAAAAAACAGAACCAATTAAT	2520
Db	2461	AAAGTATCTGTTACTGGAAGTTAGCACCTTAGGGAAGGCAAAAAACAGAACCAATTAAT	2520
OY	2521	GTTGTAGTCACTGTGTGACACATTTTGA AAAACCCCAAGGACTTAATCATGTGTTTCCAAAG	2580
Db	2521	GTTGTAGTCACTGTGTGACACATTTTGA AAAACCCCAAGGACTTAATCATGTGTTTCCAAAG	2580
OY	2581	ATAATAGAAATGACACAGAGGCTTTAGTATTCCTTTGGGACATGAATTAACCAAGCTC	2640
Db	2581	ATAATAGAAATGACACAGAGGCTTTAGTATTCCTTTGGGACATGAATTAACCAAGCTC	2640
OY	2641	GGGAACAACGATAGAAATGGAAGAAAGTGAATCTGATCTCACTATTTTGCAGATACAT	2700

Db	2641	GGGAAACAGCATTACAAATGCAAGAAAGTGAACCTTGATGCTCAGTATTGGCAGAAATACAT	2700
Qy	2701	TCAAAGTTTCAAAAGCCAGTCATTGGCTGTGTTTTCAATTCAGAAATGCAGAGAGG	2760
Db	2701	TCAAAGTTTCAAAAGCCAGTCATTGGCTGTGTTTTCAATTCAGAAATGCAGAGAGG	2760
Qy	2761	AATGTCGAACATTTCCTCCGCCCTCTGGGTCTTAAAGAAACAAAGTCCAAAAGTCACCTT	2820
Db	2761	AATGTCGAACATTTCCTCCGCCCTCTGGGTCTTAAAGAAACAAAGTCCAAAAGTCACCTT	2820
Qy	2821	TTGAATGTGAACAAAGAGAGAAATCAAGAAAGATGACTATATATCAAGCCTGTAC	2880
Db	2821	TTGAATGTGAACAAAGAGAGAAATCAAGAAAGATGAGTCTATATATCAAGCCTGTAC	2880
Qy	2881	AGACAGTTAATATCTCCTGAGGCTTTCTGTGGTGGTCAGAAAGATTAAGCAGTTGATA	2940
Db	2881	AGACAGTTAATATCTCCTGAGGCTTTCTGTGGTGGTCAGAAAGATTAAGCAGTTGATA	2940
Qy	2941	ATGCCAAATGTAGTATCAAAAGAGGCTTAGGTTTTGTCTATCATCTCAGTTCAGAGAGCA	3000
Db	2941	ATGCCAAATGTAGTATCAAAAGAGGCTTAGGTTTTGTCTATCATCTCAGTTCAGAGAGCA	3000
Qy	3001	ACGAATCTGGACTCATTTACTCCAATTAACATGACCTTTACAAACCCTATGCTATAC	3060
Db	3001	ACGAATCTGGACTCATTTACTCCAATTAACATGACCTTTACAAACCCTATGCTATAC	3060
Qy	3061	CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTATAGAAAAATCGCTAGAGG	3120
Db	3061	CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTATAGAAAAATCGCTAGAGG	3120
Qy	3121	AAAACTTTGAGAGACATTCATATGTCACCTGAAGAGAAATGGAAATGAGAACATTTCCAA	3180
Db	3121	AAAACTTTGAGAGACATTCATATGTCACCTGAAGAGAAATGGAAATGAGAACATTTCCAA	3180
Qy	3181	GTACGCTAGACACATTTAGCCGTATTAACATTAGAGAAATGTTTTTAAGAGAGCCACTT	3240
Db	3181	GTACGCTAGACACATTTAGCCGTATTAACATTAGAGAAATGTTTTTAAGAGAGCCACTT	3240
Qy	3241	CAAGCAATTAATTAATGAAGTAGTTCACCTACTAATGAAGTGGGCTCCAGTATTAATGAAG	3300
Db	3241	CAAGCAATTAATTAATGAAGTAGTTCACCTACTAATGAAGTGGGCTCCAGTATTAATGAAG	3300
Qy	3301	TAGGTTCCAGTATGATAAAAACATTCACAGACATTAAGTAGAAACAGAGGCCCAAAATTGA	3360
Db	3301	TAGGTTCCAGTATGATAAAAACATTCACAGACATTAAGTAGAAACAGAGGCCCAAAATTGA	3360
Qy	3361	ATGCTATCTTAGATTAGGGGTTTTGCAACCTGAGAGTCTATTAACCAAGTCTTCTCTGGAA	3420
Db	3361	ATGCTATCTTAGATTAGGGGTTTTGCAACCTGAGAGTCTATTAACCAAGTCTTCTCTGGAA	3420
Qy	3421	GTAAATGTAAAGATTCCTGAATAAATAAAGCAAGAATATGAAGAGTAGTTCAGACTGTTA	3480
Db	3421	GTAAATGTAAAGATTCCTGAATAAATAAAGCAAGAATATGAAGAGTAGTTCAGACTGTTA	3480
Qy	3481	ATACAGATTTTCCTCCATATCTGATTTAGATTAACCTTAGAACAGCTTAATGGGAAGTAGTC	3540
Db	3481	ATACAGATTTTCCTCCATATCTGATTTAGATTAACCTTAGAACAGCTTAATGGGAAGTAGTC	3540
Qy	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGCTGAAATTAAGG	3600
Db	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGCTGAAATTAAGG	3600
Qy	3601	AAAGTACTAGTTTGTCTGTAATAATGACATTAAGAAAGTTCGTGCTTTTTAGCAAAAGCG	3660
Db	3601	AAAGTACTAGTTTGTCTGTAATAATGACATTAAGAAAGTTCGTGCTTTTTAGCAAAAGCG	3660
Qy	3661	TCCAGAAAGAGACCTTAGCAGAGAGTCTACCCCTTACACCATTACATTTGGCTCAGG	3720
Db	3661	TCCAGAAAGAGACCTTAGCAGAGAGTCTACCCCTTACACCATTACATTTGGCTCAGG	3720
Qy	3721	GTTAACGGAAGAGGGCCCAAGAAATTTAGAGTCTTCAGAGAGACTTATCTAGTAGAGATG	3780

Db 3721 GTTACGAGAGAGGGCCAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTAGAGATG 3780  
OY 3781 AAGAGCTTCCTGCTTCCAACTTGTATTTGGTAAGTAAACATATACCTTCAGT 3840  
OY 3781 AAGAGCTTCCTGCTTCCAACTTGTATTTGGTAAGTAAACATATACCTTCAGT 3840  
Db 3841 CTACTAGAGCATGACCCGTTGGTACGAGTGTCTGTAAAGAACACAGAGGAAATTTAT 3900  
OY 3841 CTACTAGAGCATGACCCGTTGGTACGAGTGTCTGTAAAGAACACAGAGGAAATTTAT 3900  
OY 3901 TATCATTTGAGAAATAGCTTAAATGACTGAGTACAGAGTAAATTTGGCAAGGATCTC 3960  
Db 3901 TATCATTTGAGAAATAGCTTAAATGACTGAGTACAGAGTAAATTTGGCAAGGATCTC 3960  
OY 3961 AGGACATACCTTGTAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4020  
Db 3961 AGGACATACCTTGTAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4020  
OY 4021 GTGAATTTGGAAGACTTGTAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4080  
Db 4021 GTGAATTTGGAAGACTTGTAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4080  
OY 4081 CCAAAACAAATGAGGACTGAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4140  
Db 4081 CCAAAACAAATGAGGACTGAGTGAAGAACAAATGTTCTGCTAGCTTGTCTTCACTGCA 4140  
OY 4141 TTTACATGATGAGAAAGAGGAGGAGGCTTGGAAAGAAATATATCAAGAAAGCAATGG 4200  
Db 4141 TTTACATGATGAGAAAGAGGAGGAGGCTTGGAAAGAAATATATCAAGAAAGCAATGG 4200  
OY 4201 TGGATTTCAAACTTAGTGAAGAGCAGATCTGGGTGTGAGAGTGAAGAACAGCTCTGGA 4260  
Db 4201 TGGATTTCAAACTTAGTGAAGAGCAGATCTGGGTGTGAGAGTGAAGAACAGCTCTGGA 4260  
OY 4261 ACTGCTCAGGCGTATCTCTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4320  
Db 4261 ACTGCTCAGGCGTATCTCTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4320  
OY 4321 AACATTAACCTGATTAAGGCTCCAGAGGAAATGCTAAGTGAAGTGAAGTGAAGTGAAG 4380  
Db 4321 AACATTAACCTGATTAAGGCTCCAGAGGAAATGCTAAGTGAAGTGAAGTGAAGTGAAG 4380  
OY 4381 ATGGAGAGCAGCTTCTTAACAGTACCTTCCATCAATGATGATGATGATGATGATGATG 4440  
Db 4381 ATGGAGAGCAGCTTCTTAACAGTACCTTCCATCAATGATGATGATGATGATGATGATG 4440  
OY 4441 ACCTGGGAATCCAGAGAACAGACATCAAGAAAGCAATTAATTAATTAATTAATTAAT 4500  
Db 4441 ACCTGGGAATCCAGAGAACAGACATCAAGAAAGCAATTAATTAATTAATTAATTAAT 4500  
OY 4501 GTGAATTAACCTTAAGGCTCAGAGGCTTCTGCTGCAAGTGAAGTGAAGTGAAGTGA 4560  
Db 4501 GTGAATTAACCTTAAGGCTCAGAGGCTTCTGCTGCAAGTGAAGTGAAGTGAAGTGAAG 4560  
OY 4561 CAGATAGTCTCAACAGTAAATTAAGAACAGAGGAGTGAAGGATCATCCCTTCTTAAT 4620  
Db 4561 CAGATAGTCTCAACAGTAAATTAAGAACAGAGGAGTGAAGGATCATCCCTTCTTAAT 4620  
OY 4621 GGCATCATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4680  
Db 4621 GGCATCATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4680  
OY 4681 ACTACCATCTCAGAGAGGCTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4740  
Db 4681 ACTACCATCTCAGAGAGGCTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4740  
OY 4741 AGTCTGGGCGACAGATTTGAGGAAATCTTACTTGGCAAGGCAAGATCTAGAGGAA 4800  
Db 4741 AGTCTGGGCGACAGATTTGAGGAAATCTTACTTGGCAAGGCAAGATCTAGAGGAA 4800  
OY 4801 CCCCTTACCTGGAATCTGGAATCAGGCTTCTCTGATGATGATGATGATGATGATGATG 4860  
Db 4801 CCCCTTACCTGGAATCTGGAATCAGGCTTCTCTGATGATGATGATGATGATGATGATG 4860

OY 4861 AAGACAGAGCCAGAGTACGCTGTTGGCAACATATACCTTCAACCTTGCATGTA 4920  
Db 4861 AAGACAGAGCCAGAGTACGCTGTTGGCAACATATACCTTCAACCTTGCATGTA 4920  
OY 4921 AAGTCCCAATTTGAAAGTTGAGAAATCTGCCAGAGTCCAGCTGCTCATACTG 4980  
Db 4921 AAGTCCCAATTTGAAAGTTGAGAAATCTGCCAGAGTCCAGCTGCTCATACTG 4980  
OY 4981 ATACTCTGGGTATATGCAATGGAAGAAAGTGTGAGGAGGAGAAAGCCGAATTTGAC 5040  
Db 4981 ATACTCTGGGTATATGCAATGGAAGAAAGTGTGAGGAGGAGAAAGCCGAATTTGAC 5040  
OY 5041 CTTCACAGAAAGGCTCAACAAAGATGTCATGTTGTGCTGCTGACCCAGAA 5100  
Db 5041 CTTCACAGAAAGGCTCAACAAAGATGTCATGTTGTGCTGCTGACCCAGAA 5100  
OY 5101 AATTTATGCTGCTGATGAGTGTGAGAAACACACATCACTTAATCTAATTA 5160  
Db 5101 AATTTATGCTGCTGATGAGTGTGAGAAACACACATCACTTAATCTAATTA 5160  
OY 5161 CTGAGAGACTACTCATGTTGTTATGAAAGAGATGCTGATGTTGTGTGAGAGGAC 5220  
Db 5161 CTGAGAGACTACTCATGTTGTTATGAAAGAGATGCTGATGTTGTGTGAGAGGAC 5220  
OY 5221 TGAATATTTTCTAGGAATTTGCGGAGAAATGGGTAGTACTATTTCTGGGTACCC 5280  
Db 5221 TGAATATTTTCTAGGAATTTGCGGAGAAATGGGTAGTACTATTTCTGGGTACCC 5280  
OY 5281 AGCTATTTAAGAAAGAAATGCTAATGAGCATATTTTGAAGTCAAGAGAGATG 5340  
Db 5281 AGCTATTTAAGAAAGAAATGCTAATGAGCATATTTTGAAGTCAAGAGAGATG 5340  
OY 5341 TCATGGAAGAAACCCCAAGGTCGAAGGAGCAAGAGATGTCAGAGAGAAATG 5400  
Db 5341 TCATGGAAGAAACCCCAAGGTCGAAGGAGCAAGAGATGTCAGAGAGAAATG 5400  
OY 5401 TCAGGGGGTAGAAATCTGTTGATGAGGCTTCAACAAATGAGGAGATCACTG 5460  
Db 5401 TCAGGGGGTAGAAATCTGTTGATGAGGCTTCAACAAATGAGGAGATCACTG 5460  
OY 5461 AATGATGATGACGCTGTGTGCTTCTGTGTGAGAGGCTTATCATTCACCTT 5520  
Db 5461 AATGATGATGACGCTGTGTGCTTCTGTGTGAGAGGCTTATCATTCACCTT 5520  
OY 5521 GCACAGGTGTCACCAATTTGTGTGTGAGGCAATGCTGAGAGAGCAATGCT 5580  
Db 5521 GCACAGGTGTCACCAATTTGTGTGTGAGGCAATGCTGAGAGAGCAATGCT 5580  
OY 5581 TCCATCAATTTGGGAGATGTCAGAGCCTGTTGTCACCCAGAGTGGTTGGACA 5640  
Db 5581 TCCATCAATTTGGGAGATGTCAGAGCCTGTTGTCACCCAGAGTGGTTGGACA 5640  
OY 5641 GTGTACACTCTACAGGCTCAGAGCTGAGACCTGATATCCCAAGATCCCA 5700  
Db 5641 GTGTACACTCTACAGGCTCAGAGCTGAGACCTGATATCCCAAGATCCCA 5700  
OY 5701 GGCATCACTGA 5711  
Db 5701 GGCATCACTGA 5711

RESULT 3  
AAT17493  
ID AAT17493 standard; cDNA: 5914 BP.  
AC AAT17493;  
XX 02-OCT-1996 (first entry)  
XX 02-OCT-1996 (first entry)  
XX Mutated BRCA1 coding sequence from PM06.  
XX  
XX  
KW Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;



```
Db 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAAGAAATGAAATGCAAGAAACTGCCATGCT 1200
Oy 1201 CAGAGAAATCCTAGATACCTAGAGATGTTCTTGGATAACTAAATAGCAGCATTCAGA 1260
Db 1201 CAGAGAAATCCTAGATACCTAGAGATGTTCTTGGATAACTAAATAGCAGCATTCAGA 1260
Oy 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTGAAGCTGTAGTGTCTGATGACTCAGATGATG 1320
Db 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTGAAGCTGTAGTGTCTGATGACTCAGATGATG 1320
Oy 1321 GGGAGTCTGATCAAAATGCCAAAGTACGTGATGATGAGAGCTTCTAAATGAGTGAATG 1380
Db 1321 GGGAGTCTGATCAAAATGCCAAAGTACGTGATGATGAGAGCTTCTAAATGAGTGAATG 1380
Oy 1381 AATATCTGCTTCTCAGAGAAATAGACTTACTGGCCAGTATGCTCCTGAGGGCTTTA 1440
Db 1381 AATATCTGCTTCTCAGAGAAATAGACTTACTGGCCAGTATGCTCCTGAGGGCTTTA 1440
Oy 1441 TATGTAAAAGTGAAGAGTTCACTCCAAATCAGTAGAGTAATATTTGAAGCAAAATAT 1500
Db 1441 TATGTAAAAGTGAAGAGTTCACTCCAAATCAGTAGAGTAATATTTGAAGCAAAATAT 1500
Oy 1501 TTGGGAAAACCTATGGGAAAGGCAAGGCTCCCACTTAAGCCATGTAAGTGAATG 1560
Db 1501 TTGGGAAAACCTATGGGAAAGGCAAGGCTCCCACTTAAGCCATGTAAGTGAATG 1560
Oy 1561 TAATATAGAGCATTGTTTACTGAGCCAGATATATACAGAGGCTCCCTCACAATA 1620
Db 1561 TAATATAGAGCATTGTTTACTGAGCCAGATATATACAGAGGCTCCCTCACAATA 1620
Oy 1621 AATTAAGGCTAAAAGAGACCTAGATCAGGCTTCACTCTGAGATTTTATCAAGAAAG 1680
Db 1621 AATTAAGGCTAAAAGAGACCTAGATCAGGCTTCACTCTGAGATTTTATCAAGAAAG 1680
Oy 1681 CAGATTTGGCAGTCAAAAAGACTCCTGAAATGATATATGAGGAACCTAACCAAGGAGC 1740
Db 1681 CAGATTTGGCAGTCAAAAAGACTCCTGAAATGATATATGAGGAACCTAACCAAGGAGC 1740
Oy 1741 AGAATGCTCAATGATGATATTTACTAATAGTGTGATGAGATTAATAACAAAGGTGAT 1800
Db 1741 AGAATGCTCAATGATGATATTTACTAATAGTGTGATGAGATTAATAACAAAGGTGAT 1800
Oy 1801 CTATTCAGATGAGAAAAATCTTAACCCAAATAGAAATCACTCGAAAAAGATCTGCTTCA 1860
Db 1801 CTATTCAGATGAGAAAAATCTTAACCCAAATAGAAATCACTCGAAAAAGATCTGCTTCA 1860
Oy 1861 AAACGAAAGCTGAACTATTAAGCAGCAGTATTAAGCAATGTAAGTGAAGTGAATATTC 1920
Db 1861 AAACGAAAGCTGAACTATTAAGCAGCAGTATTAAGCAATGTAAGTGAAGTGAATATTC 1920
Oy 1921 ACAATTTCAAAAGCAGCTTAATAAAGATAGGCTGAGAGGAAGTCTTCTACAGGCAATTC 1980
Db 1921 ACAATTTCAAAAGCAGCTTAATAAAGATAGGCTGAGAGGAAGTCTTCTACAGGCAATTC 1980
Oy 1981 ATGCGCTTGAATAGTACTAGTAGAAATCTTAAGCCACCTAATTTGTAATTTGCAAA 2040
Db 1981 ATGCGCTTGAATAGTACTAGTAGAAATCTTAAGCCACCTAATTTGTAATTTGCAAA 2040
Oy 2041 TTGATAGTGTCTTCTAGCAGTGAAGATTAAGAAAAAAAGTACAAACCAATGCCAGTCA 2100
Db 2041 TTGATAGTGTCTTCTAGCAGTGAAGATTAAGAAAAAAAGTACAAACCAATGCCAGTCA 2100
Oy 2101 GGCACAGCAGAACTACAACTCATGGAAGTAAAGAACCTGCACTGAGGCAAGAA 2160
Db 2101 GGCACAGCAGAACTACAACTCATGGAAGTAAAGAACCTGCACTGAGGCAAGAA 2160
Oy 2161 GTTAAACAGCCAAATGAAAGACAGACAGTAAAGACATGACAGTATCTTTCCAGAGCTGA 2220
Db 2161 GTTAAACAGCCAAATGAAAGACAGACAGTAAAGACATGACAGTATCTTTCCAGAGCTGA 2220
Oy 2221 AGTTAAACAATGACACTGCTGTTTCTTACTAAGTGTTCAAATACAGAGTGAATTAAGAT 2280
Db 2221 AGTTAAACAATGACACTGCTGTTTCTTACTAAGTGTTCAAATACAGAGTGAATTAAGAT 2280
Oy 2281 TTGTCAATCCTAGCTGCTTCCAGAGAAAGAAAGAGAACTAGAAACAGTTAAAGTGT 2340
Db 2281 TTGTCAATCCTAGCTGCTTCCAGAGAAAGAAAGAGAACTAGAAACAGTTAAAGTGT 2340
Oy 2341 CTAAATAGTCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAGAAAGGTTTTGCAATCG 2400
Db 2341 CTAAATAGTCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAGAAAGGTTTTGCAATCG 2400
Oy 2401 AAAAGATCTGTAGAGAGTACATATTTCTTTGGTACTGTGACTGATTTATGACACTCAG 2460
Db 2401 AAAAGATCTGTAGAGAGTACATATTTCTTTGGTACTGTGACTGATTTATGACACTCAG 2460
Oy 2461 AAAGATCTGCTTACTGGAAGTTAGCACCTCTGAGAGAGCAAAACAGAACTAAAT 2520
Db 2461 AAAGATCTGCTTACTGGAAGTTAGCACCTCTGAGAGAGCAAAACAGAACTAAAT 2520
Oy 2521 GTGTAGTCACTGTGAGAGATTTGAAGAACCCCAAGGACTAATGATGTTTCCAAAG 2580
Db 2521 GTGTAGTCACTGTGAGAGATTTGAAGAACCCCAAGGACTAATGATGTTTCCAAAG 2580
Oy 2581 ATTAATGAATGACACAGAGGCTTTAAGTATCCATTTGGGACATGAGTTAACACAGTC 2640
Db 2581 ATTAATGAATGACACAGAGGCTTTAAGTATCCATTTGGGACATGAGTTAACACAGTC 2640
Oy 2641 GGGAAACAAGCATAGAAATGAGAAAGAGTGAAGTCTGATGATCTGATTTGCAAGATACAT 2700
Db 2641 GGGAAACAAGCATAGAAATGAGAAAGAGTGAAGTCTGATGATCTGATTTGCAAGATACAT 2700
Oy 2701 TCAAGGTTTCAAGGCGCCAGCATTTGCTCTGTTTCAATCCAGAAATGCAAGAGAG 2760
Db 2701 TCAAGGTTTCAAGGCGCCAGCATTTGCTCTGTTTCAATCCAGAAATGCAAGAGAG 2760
Oy 2761 AATGTCCAACTTCTCTGCCCCTCTGAGGCTCTTAAAGAAACAAAGTCCAAAGTCACT 2820
Db 2761 AATGTCCAACTTCTCTGCCCCTCTGAGGCTCTTAAAGAAACAAAGTCCAAAGTCACT 2820
Oy 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGAGAAAGATAGTCTAATATCAAGCCGTAC 2880
Db 2821 TTGAATGTGAACAAAGAGAAAGAAATCAAGAGAAAGATAGTCTAATATCAAGCCGTAC 2880
Oy 2881 AGACAGTTAATATCACTGAGGCTTCTGCTGTGTGTCAGAAAGATTAAGCAGTTGATA 2940
Db 2881 AGACAGTTAATATCACTGAGGCTTCTGCTGTGTGTCAGAAAGATTAAGCAGTTGATA 2940
Oy 2941 ATGCCAAATGATGATCAAGAGAGGCTGATGTTTGTATATCTCAGTTGAGAGCA 3000
Db 2941 ATGCCAAATGATGATCAAGAGAGGCTGATGTTTGTATATCTCAGTTGAGAGCA 3000
Oy 3001 ACAGAACTGAGACTATTACTCCAAATTAACATGAGACTTTTACAAAACCCATATCGATAC 3060
Db 3001 ACAGAACTGAGACTATTACTCCAAATTAACATGAGACTTTTACAAAACCCATATCGATAC 3060
Oy 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTGAGAG 3120
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTGAGAG 3120
Oy 3121 AAAACTTTGAGAGACATTTAATGTCACTGGAAGAGAAATGGGAAATGGAACATTCCAA 3180
Db 3121 AAAACTTTGAGAGACATTTAATGTCACTGGAAGAGAAATGGGAAATGGAACATTCCAA 3180
Oy 3181 GTACAGTGAAGCAATTTAGCCCTAATTAACATTAAGAAATATGTTTAAAGAACCCAGCT 3240
Db 3181 GTACAGTGAAGCAATTTAGCCCTAATTAACATTAAGAAATATGTTTAAAGAACCCAGCT 3240
Oy 3241 CAAGCAATATTAATGAAGTAGTGTCCAGTACTAATGAAGTGGCTCCACTATTAATGAAA 3300
Db 3241 CAAGCAATATTAATGAAGTAGTGTCCAGTACTAATGAAGTGGCTCCACTATTAATGAAA 3300
Oy 3301 TAGGTTCCAGTATGAAAAATTTCAAGCAGAACTAGTGAAGAACAGAGGCCCAAAATGCA 3360
Db 3301 TAGGTTCCAGTATGAAAAATTTCAAGCAGAACTAGTGAAGAACAGAGGCCCAAAATGCA 3360
```

```
OY 3361 ATGCTATGGTTAGATTAGGGGTTTTGCACCTGAGGCTCTATAAACAAGCTCTTCGGAA 3420
|||||
DB 3361 ATGCTATGGTTAGATTAGGGGTTTTGCACCTGAGGCTCTATAAACAAGCTCTTCGGAA 3420
OY 3421 GTAATTGTAGCATCTCTGAATAAATAAAGCAAGATATAGAGAGTGTTCAGACTGTTA 3480
|||||
DB 3421 GTAATTGTAGCATCTCTGAATAAATAAAGCAAGATATAGAGAGTGTTCAGACTGTTA 3480
OY 3481 ATACAGATTTCTCTCATATCTGATTTGATTAAGTACCTATAGACAGCTATGGGAGTATGTC 3540
|||||
DB 3481 ATACAGATTTCTCTCATATCTGATTTGATTAAGTACCTATAGACAGCTATGGGAGTATGTC 3540
OY 3541 ATGATCTCAGAGTTTGTCTGAGACACCTGATGACCTGTAGATGATGTGAATAAAGG 3600
|||||
DB 3541 ATGATCTCAGAGTTTGTCTGAGACACCTGATGACCTGTAGATGATGTGAATAAAGG 3600
OY 3601 AAGATCTAGTTTCTCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTACCAAAAGCG 3660
|||||
DB 3601 AAGATCTAGTTTCTCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTACCAAAAGCG 3660
OY 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCCAGGCCCTTTCACCATACACATTTGGCTCAGG 3720
|||||
DB 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCCAGGCCCTTTCACCATACACATTTGGCTCAGG 3720
OY 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCCAGGCCCTTTCACCATACACATTTGGCTCAGG 3720
|||||
DB 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCCAGGCCCTTTCACCATACACATTTGGCTCAGG 3720
OY 3721 GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGTGAGATG 3780
|||||
DB 3721 GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGTGAGATG 3780
OY 3781 AAGAGCTTCCCTGCTTCACACCTGTTATTTGGTAAGTAAACAATATACCTTCACAT 3840
|||||
DB 3781 AAGAGCTTCCCTGCTTCACACCTGTTATTTGGTAAGTAAACAATATACCTTCACAT 3840
OY 3841 CTACTAGGCATAGACACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTAT 3900
|||||
DB 3841 CTACTAGGCATAGACACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTAT 3900
OY 3901 TATATTGAAGAAATAGCTTAATATGCTGACAGTACACAGGTAATATTGGCAAGGCAATCTC 3960
|||||
DB 3901 TATATTGAAGAAATAGCTTAATATGCTGACAGTACACAGGTAATATTGGCAAGGCAATCTC 3960
OY 3961 AGGAACATCACCTTAGTAGAGGAAACAATGTTGCTGAGTGTGTTCTTACAGGTGA 4020
|||||
DB 3961 AGGAACATCACCTTAGTAGAGGAAACAATGTTGCTGAGTGTGTTCTTACAGGTGA 4020
OY 4021 GTGAATTGGAAGACTTGACTCAATATACAAACACCCAGAGATCCTTCTTGATTGGTCTT 4080
|||||
DB 4021 GTGAATTGGAAGACTTGACTCAATATACAAACACCCAGAGATCCTTCTTGATTGGTCTT 4080
OY 4081 CCAACAACATAGAGCATGCTGTAAGACCCAGGGAGTTGCTGAGTGACAAAGAAATTGG 4140
|||||
DB 4081 CCAACAACATAGAGCATGCTGTAAGACCCAGGGAGTTGCTGAGTGACAAAGAAATTGG 4140
OY 4141 TTTGAGATGATGAGAAAGAGAGAGCGGCTTGAAGAAATATATAGAAAGAGCAAGCA 4200
|||||
DB 4141 TTTGAGATGATGAGAAAGAGAGAGCGGCTTGAAGAAATATATAGAAAGAGCAAGCA 4200
OY 4201 TGAATTCAAACTTAGTGAGAGCAGCATCTGGGTGAGAGTGAACAAGCGCTCTGAAG 4260
|||||
DB 4201 TGAATTCAAACTTAGTGAGAGCAGCATCTGGGTGAGAGTGAACAAGCGCTCTGAAG 4260
OY 4261 ACTGCTCAGGGCTATCTCTCAGAGTGAATTTTACCACTCAGAGAGGGATACCATG 4320
|||||
DB 4261 ACTGCTCAGGGCTATCTCTCAGAGTGAATTTTACCACTCAGAGAGGGATACCATG 4320
OY 4321 AACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAAGTGAAGCTGTGTAGAACAGC 4380
|||||
DB 4321 AACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAAGTGAAGCTGTGTAGAACAGC 4380
OY 4381 ATGGAGCGACGCTTCTTAACAGCTAACCTTCATCATTAAGTACTCTCTGCCCTTGAAG 4440
|||||
DB 4381 ATGGAGCGACGCTTCTTAACAGCTAACCTTCATCATTAAGTACTCTCTGCCCTTGAAG 4440
OY 4441 ACCTGCGAATCCAGAAACAAAGCAATCAGAAAAAGCAGTATTAACTTCACAGAAAAAGTA 4500
|||||
```

Db 5521 GCACAGCTGTCACCCCAATTTGTTGTGACGACCATGCTGACAGACAGACAAATGGCT 5580  
 Qy 5581 TCCATGCAATTTGGCAGATGTGTGAGCAGCCTGTGTGACCCGAGAGTGGTGTGACA 5640  
 Db 5581 TCCATGCAATTTGGCAGATGTGTGAGCAGCCTGTGTGACCCGAGAGTGGTGTGACA 5640  
 Qy 5641 GTGAGCACTTACCAGTGGCAGAGCTGACACCTTACCTGATACCCAGATCCCCACA 5700  
 Db 5641 GTGAGCACTTACCAGTGGCAGAGCTGACACCTTACCTGATACCCAGATCCCCACA 5700  
 Qy 5701 GCCACTACTGA 5711  
 Db 5701 GCCACTACTGA 5711

RESULT 4  
 AAT17495  
 ID AAT17495 standard; cDNA: 5914 BP.  
 XX AAT17495;  
 AC AAT17495;  
 DT 02-OCT-1996 (first entry)  
 XX  
 DE Mutated BRCA1 coding sequence from PM09.  
 XX  
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;  
 KW antibody production; germline alteration; probe; lesion neoplasia; human;  
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX CDS 120..5711  
 FT /\*tag= a  
 FT /product= BRCA1 mutant  
 FT 2201  
 FT /\*tag= b  
 FT mutation  
 FT /\*note= "C to T mutation"  
 XX  
 XX MO9605306-A2.  
 PD 22-FEB-1996.  
 XX  
 PE 11-AUG-1995; 95WO-US10202.  
 XX  
 PR 07-JUN-1995; 95US-0483553.  
 PR 12-AUG-1994; 94US-0289221.  
 PR 02-SEP-1994; 94US-0300266.  
 PR 16-SEP-1994; 94US-0308104.  
 PR 29-NOV-1994; 94US-0348824.  
 PR 24-MAR-1995; 95US-0409305.  
 PR 07-JUN-1995; 95US-0480784.  
 XX  
 PA (MIRI-) MYRIAD GENETICS INC.  
 PA (CANC-) CANCER INST.  
 PA (RECH-) CENT RECH DU CHUL.  
 PI  
 PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DK.  
 DR WPI; 1996-139702/14.  
 DR P-PSDB; AAR81481.  
 XX  
 XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers  
 XX  
 XX Claim 1; : 218pp: English.  
 XX  
 CC AAT17439-T17453 and AAT17455-T17529 represent mutations of the human  
 CC breast and ovarian cancer predisposing gene (BRCA1) (for wild type see  
 CC AAT17438). Proteins encoded by these mutations (see AAR81483-R81487 and  
 CC AAR81499-R81546) can be used as immunogens for antibody production.  
 CC These mutant BRCA1 genes have at least 1 mutation or polymorphism in

CC comparison to the wild type sequence. By detecting a germline alteration  
 CC in the wild type BRCA1 gene, a predisposition for breast and ovarian  
 CC cancer can be diagnosed. In one method, BRCA1 mRNA isolated from a tissue  
 CC sample from a subject has a probe, corresponding to a fragment of the  
 CC wild type sequence (or an allele-specific probe for one of these  
 CC mutations), added to it. The conditions allow for hybridisation of the  
 CC probe to the mRNA, and any hybridisation which occurs is detected.  
 CC Alternatively the BRCA1 gene in the tissue sample is isolated, and a  
 CC shift in electrophoretic mobility of single stranded DNA from the sample  
 CC on a non-denaturing polyacrylamide gel indicates a mutation. These  
 CC methods of detection can also diagnose a lesion neoplasia associated with  
 CC the BRCA1 locus. The methods may be used in gene therapy, protein  
 CC replacement therapy and protein mimetics, and may be used to screen for  
 CC drugs in cancer therapy.  
 XX  
 S0 Sequence 5914 BP: 2006 A; 1155 C; 1316 G; 1437 T; 0 other;

Query Match 100.0%; Score 5709.4; DB 17; Length 5914;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCTGCGTGAGACCTTCCGTGACCCCGACAGGCTGTGGGTTCTCAGATACTGGCC 60  
 Db 1 AGCTGCGTGAGACCTTCCGTGACCCCGACAGGCTGTGGGTTCTCAGATACTGGCC 60  
 Qy 61 CCTGCGCTCAGAGAGCCCTTACACCTCTGCTGGGTAAAGTTCAATGGAACAGAAAGAA 120  
 Db 61 CCTGCGCTCAGAGAGCCCTTACACCTCTGCTGGGTAAAGTTCAATGGAACAGAAAGAA 120  
 Qy 121 TGGATTATATCTGCTCTTCCGCTGAGAGATACAAATGTCTAATGCTATGAGAAA 180  
 Db 121 TGGATTATATCTGCTCTTCCGCTGAGAGATACAAATGTCTAATGCTATGAGAAA 180  
 Qy 181 TCTTAGAGTGTCCATCTGTGTGAGTGTATCAAGAAAGTCTCTCCAAAGTGTGACC 240  
 Db 181 TCTTAGAGTGTCCATCTGTGTGAGTGTATCAAGAAAGTCTCTCCAAAGTGTGACC 240  
 Qy 241 ACATATTTTGGCAATTTTGCATGCTGAAACTTCTCAACACAGAAAGGGCTTCACAGT 300  
 Db 241 ACATATTTTGGCAATTTTGCATGCTGAAACTTCTCAACACAGAAAGGGCTTCACAGT 300  
 Qy 301 GTCCCTTATATGATATGATATTAACCAAAAGAGCCTACAGAAAGTACGATTTAGTC 360  
 Db 301 GTCCCTTATATGATATGATATTAACCAAAAGAGCCTACAGAAAGTACGATTTAGTC 360  
 Qy 361 AACTTTGTTGAAGACCTATGAAATCATTTGTGCTTTGACCTTGACACAGGTTGGAGT 420  
 Db 361 AACTTTGTTGAAGACCTATGAAATCATTTGTGCTTTGACCTTGACACAGGTTGGAGT 420  
 Qy 421 ATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCCGACATCTAAAGATG 480  
 Db 421 ATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCCGACATCTAAAGATG 480  
 Qy 481 AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAGACTTTACAGAGTG 540  
 Db 481 AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAGACTTTACAGAGTG 540  
 Qy 541 AACCCGAAATCCTTCTCTGAGAGAAACAGCTCAGTGTCTCAACTCTTAACCTGGAA 600  
 Db 541 AACCCGAAATCCTTCTCTGAGAGAAACAGCTCAGTGTCTCAACTCTTAACCTGGAA 600  
 Qy 601 CTGTGAGAACTCTGAGAGCAAGAGGAGATACCACTCAAAAGACGCTGTCTACATTTG 660  
 Db 601 CTGTGAGAACTCTGAGAGCAAGAGGAGATACCACTCAAAAGACGCTGTCTACATTTG 660  
 Qy 661 AATTGGATCTGATTTCTTGTGAAGATACCGTTAATAGGCAACTTATGGAGTGTGGAG 720  
 Db 661 AATTGGATCTGATTTCTTGTGAAGATACCGTTAATAGGCAACTTATGGAGTGTGGAG 720  
 Qy 721 ATCAAGATTTGTTCAAATATCAACCCCTCAAGAGACAGGATGAATAGTTGGATTCTG 780  
 Db 721 ATCAAGATTTGTTCAAATATCAACCCCTCAAGAGACAGGATGAATAGTTGGATTCTG 780



QY 781 CAAAAAGGCTGCTGTGAAATTTTCTGAGCGGATGTAAACAATACTGAACATCATCAAC 840  
DB 781 CAAAAAGGCTGCTGTGAAATTTTCTGAGCGGATGTAAACAATACTGAACATCATCAAC 840  
QY 841 CCAATTAATATGATTTTGAACACCACTGAGAAAGCTGACGCTGAGAGGCTTCCACAAAAGT 900  
DB 841 CCAATTAATATGATTTTGAACACCACTGAGAAAGCTGACGCTGAGAGGCTTCCACAAAAGT 900  
QY 901 ATGAGGTAATGCTGTTTCAAACTTGCATGTGAGCCATGTGGACAAATCTCATGCA 960  
DB 901 ATGAGGTAATGCTGTTTCAAACTTGCATGTGAGCCATGTGGACAAATCTCATGCA 960  
QY 961 GCTCATATACAGCATGAGAACAGCAGTTTATTACTCATAAAGACAGATGATAGAAA 1020  
DB 961 GCTCATATACAGCATGAGAACAGCAGTTTATTACTCATAAAGACAGATGATAGAAA 1020  
QY 1021 AGGCTGAATCTGTAAATTAAGCAACAGCTGGCTTACGAGAGGCAACATTAACGAT 1080  
DB 1021 AGGCTGAATCTGTAAATTAAGCAACAGCTGGCTTACGAGAGGCAACATTAACGAT 1080  
QY 1081 GGGCTGGAAGTGAAGAAACATGTATGATAGGCGGACTCCAGCAGCAAAAAAGTAG 1140  
DB 1081 GGGCTGGAAGTGAAGAAACATGTATGATAGGCGGACTCCAGCAGCAAAAAAGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAAAAAGATGGAATAGCAGAAACTGCCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGTGAGAAAAAGATGGAATAGCAGAAACTGCCATGCT 1200  
QY 1201 CAGAGATCTGTAGAGATGCTGAGAGATGCTTCTGTGATTAACATTAATAGCAGCTTACA 1260  
DB 1201 CAGAGATCTGTAGAGATGCTGAGAGATGCTTCTGTGATTAACATTAATAGCAGCTTACA 1260  
QY 1261 AAGTTAATGAGTGTTCAGAAAGTATGATGATGATGATGATGATGATGATGATGATG 1320  
DB 1261 AAGTTAATGAGTGTTCAGAAAGTATGATGATGATGATGATGATGATGATGATGATG 1320  
QY 1321 GGGAGTCTGAATTAATGCCAAAGTACGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 GGGAGTCTGAATTAATGCCAAAGTACGATGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 AATATCTGCTGTTCTGAGAAATAGACTTACTGGCAGTATCCATGAGGCTTAA 1440  
DB 1381 AATATCTGCTGTTCTGAGAAATAGACTTACTGGCAGTATCCATGAGGCTTAA 1440  
QY 1441 TATGTAAGTGAAGAGTCTCACTCAAAATCAGTAGAGTAATATTGAAGACAAAATAT 1500  
DB 1441 TATGTAAGTGAAGAGTCTCACTCAAAATCAGTAGAGTAATATTGAAGACAAAATAT 1500  
QY 1501 TTGGGAAAACCTATCGGAGAGGCAAGCCTCCCACTTAAGCATGTAAGTGAATATC 1560  
DB 1501 TTGGGAAAACCTATCGGAGAGGCAAGCCTCCCACTTAAGCATGTAAGTGAATATC 1560  
QY 1561 TATATAGGAGATTTGTTACTGAGCCACAGATTAATACAAAGCGTCCCTCACAATA 1620  
DB 1561 TATATAGGAGATTTGTTACTGAGCCACAGATTAATACAAAGCGTCCCTCACAATA 1620  
QY 1621 AATTAAGGTAAGAGAGACTACATCAGGCTTCACTGAGAGATTTATACAGAAAG 1680  
DB 1621 AATTAAGGTAAGAGAGACTACATCAGGCTTCACTGAGAGATTTATACAGAAAG 1680  
QY 1681 CAGATTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAACTAACCAACGAGC 1740  
DB 1681 CAGATTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAACTAACCAACGAGC 1740  
QY 1741 AGAATGCTCAAGTGAATATTAATGATGATGATGATGATGATGATGATGATGATGAT 1800  
DB 1741 AGAATGCTCAAGTGAATATTAATGATGATGATGATGATGATGATGATGATGATGAT 1800  
QY 1801 CTATTCAGATGAGAAATCTTAACCAATAGAAATCAGCAAAAAAGATGCTTCA 1860  
DB 1801 CTATTCAGATGAGAAATCTTAACCAATAGAAATCAGCAAAAAAGATGCTTCA 1860  
QY 1861 AATGCAAACTGATTAAGCAGATTAAGCAATATGAACTCGAATTAATATCC 1920  
DB 1861 AATGCAAACTGATTAAGCAGATTAAGCAATATGAACTCGAATTAATATCC 1920  
QY 1921 ACAATTTAAAAGCACCCTAAAAGATATGCTGAGAGGAAAGTCTTACCAGCATATTC 1980  
DB 1921 ACAATTTAAAAGCACCCTAAAAGATATGCTGAGAGGAAAGTCTTACCAGCATATTC 1980  
QY 1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
DB 1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
QY 2041 TTGATAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
DB 2041 TTGATAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
QY 2101 GGCACACCAAACTTACACTGATGAGAGTAAAGACCTGCACTGAGAGCAGAGAGA 2160  
DB 2101 GGCACACCAAACTTACACTGATGAGAGTAAAGACCTGCACTGAGAGCAGAGAGA 2160  
QY 2161 GTAAACAGCCAAATGAACAGACATAAGATGAAGATGATGATGATGATGATGATGATGAT 2220  
DB 2161 GTAAACAGCCAAATGAACAGACATAAGATGAAGATGATGATGATGATGATGATGATGAT 2220  
QY 2221 AGTTAACAATGCACTGCTTTCTTACTAAGTGTCAAAATACAGTGAATTAAGAT 2280  
DB 2221 AGTTAACAATGCACTGCTTTCTTACTAAGTGTCAAAATACAGTGAATTAAGAT 2280  
QY 2281 TTGTCATCTTACCTTCCAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2340  
DB 2281 TTGTCATCTTACCTTCCAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 2340  
QY 2341 CTAAATATGCTGAAGACCCCAAGATCTATGATGATGATGATGATGATGATGATGATGAT 2400  
DB 2341 CTAAATATGCTGAAGACCCCAAGATCTATGATGATGATGATGATGATGATGATGATGAT 2400  
QY 2401 AAGATCTGTAGAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
DB 2401 AAGATCTGTAGAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
QY 2461 AAGATCTGTAGAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520  
DB 2461 AAGATCTGTAGAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520  
QY 2521 GTGTAGTCACTGATGAGATTTGAAACCCCAAGGACTAATTTCTGTTTCCAAAG 2580  
DB 2521 GTGTAGTCACTGATGAGATTTGAAACCCCAAGGACTAATTTCTGTTTCCAAAG 2580  
QY 2581 ATATATGAAATGACAGAGAGCTTTAAGTATCATTGGACATGAGTTAACACAGTC 2640  
DB 2581 ATATATGAAATGACAGAGAGCTTTAAGTATCATTGGACATGAGTTAACACAGTC 2640  
QY 2641 GGGAAACCAAGCATGAAATGAGAAAGTGAAGTATGATGATGATGATGATGATGATGAT 2700  
DB 2641 GGGAAACCAAGCATGAAATGAGAAAGTGAAGTATGATGATGATGATGATGATGATGAT 2700  
QY 2701 TCAAGGTTTCAAGCCCACTATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
DB 2701 TCAAGGTTTCAAGCCCACTATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
QY 2761 AATGTCACAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
DB 2761 AATGTCACAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
QY 2821 TTGAATGTAACAAAGGAAAGAAATCAAGGAAAGAAAGTATGATGATGATGATGATGAT 2880  
DB 2821 TTGAATGTAACAAAGGAAAGAAATCAAGGAAAGAAAGTATGATGATGATGATGATGAT 2880  
QY 2881 AGACAGTTAATATCAGTGAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
DB 2881 AGACAGTTAATATCAGTGAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
QY 2941 ATGCCAAATGATTAACAGAGGCTTGTGTTTCTATCATCTGATGATGATGATGATGATGAT 3000  
DB 2941 ATGCCAAATGATTAACAGAGGCTTGTGTTTCTATCATCTGATGATGATGATGATGATGAT 3000



D 2941 ATGCAATGTAGTATCAAGAGAGCGCTGAGTTTGTCTATCATCTCAGTTCAGAGCA 3000  
Q 3001 ACGAACTGAGCTCATCTACTCCAAATTAACATGAGACTTTTACAAACCATTATGTATAC 3060  
D 3001 ACGAACTGAGCTCATCTACTCCAAATTAACATGAGACTTTTACAAACCATTATGTATAC 3060  
Q 3061 CACCACTTTTCCATCAAGTCAATTTGTAAACTAATGTAAGAAAACTGCTATAGG 3120  
D 3061 CACCACTTTTCCATCAAGTCAATTTGTAAACTAATGTAAGAAAACTGCTATAGG 3120  
Q 3121 AAAAATTGAGAGCAATTCATCAATGCTACCTGAAGAGAAATGGAAATGAGAACATTTCAA 3180  
D 3121 AAAAATTGAGAGCAATTCATCAATGCTACCTGAAGAGAAATGGAAATGAGAACATTTCAA 3180  
Q 3181 GTACAGTGAGACAAATTAAGCCGTAATTAACATTTAGAGAAATGTTTTTAAAGAACCGAGCT 3240  
D 3181 GTACAGTGAGACAAATTAAGCCGTAATTAACATTTAGAGAAATGTTTTTAAAGAACCGAGCT 3240  
Q 3241 CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATCAA 3300  
D 3241 CAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATCAA 3300  
Q 3301 TAGGTTCCAGTATGAAGAAATTCACAGCAAGACTAGTAGAACAGAGGCGCAAAATTTGA 3360  
D 3301 TAGGTTCCAGTATGAAGAAATTCACAGCAAGACTAGTAGAACAGAGGCGCAAAATTTGA 3360  
Q 3361 ATGCTATGCTAGTATTAAGGGGTTTTCGACCTGAGGCTATAAACAAAGCTTCTCTGAA 3420  
D 3361 ATGCTATGCTAGTATTAAGGGGTTTTCGACCTGAGGCTATAAACAAAGCTTCTCTGAA 3420  
Q 3421 GTATATGTAGCATCCCGAAATTAAGAAAGCAAGAAATGAAGAGTATGTTTCAGACTTTA 3480  
D 3421 GTATATGTAGCATCCCGAAATTAAGAAAGCAAGAAATGAAGAGTATGTTTCAGACTTTA 3480  
Q 3481 ATACAGATTTCTCTCCATCTGATCTGATTCAGATTAACAGAACAGCCATGGAGATGTC 3540  
D 3481 ATACAGATTTCTCTCCATCTGATCTGATTCAGATTAACAGAACAGCCATGGAGATGTC 3540  
Q 3541 ATGCAATCTAGGTTTGTTCGAGACACCTGATGACCTGTTAGATGATGTGAATAAAGG 3600  
D 3541 ATGCAATCTAGGTTTGTTCGAGACACCTGATGACCTGTTAGATGATGTGAATAAAGG 3600  
Q 3601 AAGATCTAGTATTTGCGAAATGACATTAAGAGAACTTCTGCTTTTACGAAAGCG 3660  
D 3601 AAGATCTAGTATTTGCGAAATGACATTAAGAGAACTTCTGCTTTTACGAAAGCG 3660  
Q 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTACACCATACCATTTGGCTCAGG 3720  
D 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTACACCATACCATTTGGCTCAGG 3720  
Q 3721 GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTAGAGATG 3780  
D 3721 GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTAGAGATG 3780  
Q 3781 AAGAGCTTCCCTGCTTCCACACTTGTATTTGGTAAAGTAACATATACCTTTCAGT 3840  
D 3781 AAGAGCTTCCCTGCTTCCACACTTGTATTTGGTAAAGTAACATATACCTTTCAGT 3840  
Q 3841 CTACTAGAGCTAGCAGCGTTGCTACCGAGTGTCTGTAGAGACACAGAGAGAAATTTAT 3900  
D 3841 CTACTAGAGCTAGCAGCGTTGCTACCGAGTGTCTGTAGAGACACAGAGAGAAATTTAT 3900  
Q 3901 TATCATTTGAAGATAGCTTAATGACTGACGTACAGTAACAGATTAATTTGGCAAGGCAATCTC 3960  
D 3901 TATCATTTGAAGATAGCTTAATGACTGACGTACAGTAACAGATTAATTTGGCAAGGCAATCTC 3960  
Q 3961 AGGAACATCACCCTTAGTGAGAGAAACAATGTTCTGTAGCTTGTTCACAGTGA 4020  
D 3961 AGGAACATCACCCTTAGTGAGAGAAACAATGTTCTGTAGCTTGTTCACAGTGA 4020  
Q 4021 GTGAATTTGAAGACTTGAGTCAAAATGCAAAACACCCAGATCTTCTTATTTGGTCTT 4080  
D 4021 GTGAATTTGAAGACTTGAGTCAAAATGCAAAACACCCAGATCTTCTTATTTGGTCTT 4080

Q 4081 CCAACAAATGAGCATCAGTCTGAAAAAGCCAGAGGATTTGGTCTGAGTGACAGCAATTTGG 4140  
D 4081 CCAACAAATGAGCATCAGTCTGAAAAAGCCAGAGGATTTGGTCTGAGTGACAGCAATTTGG 4140  
Q 4141 TTTTCAGATGATGAAGAAAGAGGAGCGGCTTGGAGAAATTAATCAAGAAAGAGCAAGCA 4200  
D 4141 TTTTCAGATGATGAAGAAAGAGGAGCGGCTTGGAGAAATTAATCAAGAAAGAGCAAGCA 4200  
Q 4201 TGGATTCAAACTTAGTGGAAGCAAGCATCTGGGTGTGAGAGTGAAGCAAGCGCTCTCTAAG 4260  
D 4201 TGGATTCAAACTTAGTGGAAGCAAGCATCTGGGTGTGAGAGTGAAGCAAGCGCTCTCTAAG 4260  
Q 4261 ACTGCTCAGGGCTATCTCTCAGAGTACATTTTAACCACTCAGCAGAGGATACATGTC 4320  
D 4261 ACTGCTCAGGGCTATCTCTCAGAGTACATTTTAACCACTCAGCAGAGGATACATGTC 4320  
Q 4321 AACATACCTGATTAAGCTCAGCAGAGAAATGCTGAACTAAGAAAGCTGTGTACACAGC 4380  
D 4321 AACATACCTGATTAAGCTCAGCAGAGAAATGCTGAACTAAGAAAGCTGTGTACACAGC 4380  
Q 4381 ATGGAGCCAGCCCTTACAGAGTACCCCTTCATTAAGTGAATCTTCTGCCCCCTGAGG 4440  
D 4381 ATGGAGCCAGCCCTTACAGAGTACCCCTTTCATTAAGTGAATCTTCTGCCCCCTGAGG 4440  
Q 4441 ACCTGCGAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTTCACAGAAAAAGTA 4500  
D 4441 ACCTGCGAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTTCACAGAAAAAGTA 4500  
Q 4501 GTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCCTTCTGCTGACAACTTTGAGGTGCTG 4560  
D 4501 GTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCCTTCTGCTGACAACTTTGAGGTGCTG 4560  
Q 4561 CAGATAGTCTACAGTAAATTAAGAAACAGAGAGGAAAGTATCCCTCTTAAT 4620  
D 4561 CAGATAGTCTACAGTAAATTAAGAAACAGAGAGGAAAGTATCCCTCTTAAT 4620  
Q 4621 GCCCATCTTAGATGATAGTGTGTACATGCAAGTGTCTGAGGATCTTCAAGTAAGAA 4680  
D 4621 GCCCATCTTAGATGATAGTGTGTACATGCAAGTGTCTGAGGATCTTCAAGTAAGAA 4680  
Q 4681 ACTACCATCTCAAGAGAGAGCTCATTAAGGTGTGTGATGTGAGAGCAACAGCTGGAAG 4740  
D 4681 ACTACCATCTCAAGAGAGAGCTCATTAAGGTGTGTGATGTGAGAGCAACAGCTGGAAG 4740  
Q 4741 AGTCTGGGCGACAGATTTTACGGAACATCTTACTTCCAGAGGCAAGATCTAGAGGAA 4800  
D 4741 AGTCTGGGCGACAGATTTTACGGAACATCTTACTTCCAGAGGCAAGATCTAGAGGAA 4800  
Q 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGTATGACCCCTGAATCTGATCCTTCTG 4860  
D 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGTATGACCCCTGAATCTGATCCTTCTG 4860  
Q 4861 AAGACAGAGCCCCAGAGTCAAGCTGTGTGTGGCAACATACCATCTTCAACCTTGCATTTGA 4920  
D 4861 AAGACAGAGCCCCAGAGTCAAGCTGTGTGTGGCAACATACCATCTTCAACCTTGCATTTGA 4920  
Q 4921 AAGTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGAGTCCAGCTGCTGCTCATACTAG 4980  
D 4921 AAGTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGAGTCCAGCTGCTGCTCATACTAG 4980  
Q 4981 ATACTGCTGGGTATTAATGGAAGAAAGTGTGAGCAGAGGAGAAAGCCAGAAATTTGACAG 5040  
D 4981 ATACTGCTGGGTATTAATGGAAGAAAGTGTGAGCAGAGGAGAAAGCCAGAAATTTGACAG 5040  
Q 5041 CTTTCAACAGAAAGGCTCAACAAAAAGATGTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100  
D 5041 CTTTCAACAGAAAGGCTCAACAAAAAGATGTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100  
Q 5101 AATTTATGCTGT 5160  
D 5101 AATTTATGCTGT 5160

QY 5161 CTGAAGAGACTACTCATGTTGTTATGAAAAACAGATGCTGAGTTGTTGTGTAACGGACAC 5220  
 |||||||  
 DB 5161 CTGAAGAGACTACTCATGTTGTTATGAAAAACAGATGCTGAGTTGTTGTGTAACGGACAC 5220  
 QY 5221 TGAATAATTTTCTAGGAATTTCCGGAGAGAAATGGTAGTACCTATTTTGGGTGACCC 5280  
 |||||||  
 DB 5221 TGAATAATTTTCTAGGAATTTCCGGAGAGAAATGGTAGTACCTATTTTGGGTGACCC 5280  
 QY 5281 AGCTCTTAAAGAAAAAATGCTGAATGAGCATATTTTGAAGTCAGAGAGAGATGTGG 5340  
 |||||||  
 DB 5281 AGCTCTTAAAGAAAAAATGCTGAATGAGCATATTTTGAAGTCAGAGAGAGATGTGG 5340  
 QY 5341 TCAATGGAAGAAACCAAGAGTCCAAAGGAGACAGAAATCCAGACAGAAAGATCT 5400  
 |||||||  
 DB 5341 TCAATGGAAGAAACCAAGAGTCCAAAGGAGACAGAAATCCAGACAGAAAGATCT 5400  
 QY 5401 TCAGAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTTCACCAACATGCCCAGATCAACTGG 5460  
 |||||||  
 DB 5401 TCAGAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTTCACCAACATGCCCAGATCAACTGG 5460  
 QY 5461 AATGATGTGACAGTGTGGTGTCTTCTGTTGTAAGAGGCTTTCATCATTCACCCCTTG 5520  
 |||||||  
 DB 5461 AATGATGTGACAGTGTGGTGTCTTCTGTTGTAAGAGGCTTTCATCATTCACCCCTTG 5520  
 QY 5521 GCACAGGTGTCCACCAATTTGCTTGTGTCAGCCAGATGCTTCAGAGAGAGACAAATGGCT 5580  
 |||||||  
 DB 5521 GCACAGGTGTCCACCAATTTGCTTGTGTCAGCCAGATGCTTCAGAGAGAGACAAATGGCT 5580  
 QY 5581 TCCATGTCATTTGGGCGAGATGTGTGAGAGGCACCTGTGTGACCCGAGAGTGGGTGTGACA 5640  
 |||||||  
 DB 5581 TCCATGTCATTTGGGCGAGATGTGTGAGAGGCACCTGTGTGACCCGAGAGTGGGTGTGACA 5640  
 QY 5641 GTGTAGCACTTACAGTCCAGAGAGCTGACACTACCTGATATACCCAGATCCCCACACA 5700  
 |||||||  
 DB 5641 GTGTAGCACTTACAGTCCAGAGAGCTGACACTACCTGATATACCCAGATCCCCACACA 5700  
 QY 5701 GCCACTACTGA 5711  
 |||||||  
 DB 5701 GCCACTACTGA 5711  
 RESULT 5  
 AAV46463  
 ID AAV46463 standard; cDNA; 5711 BP.  
 AC AAV46463;  
 XX  
 DT 18-NOV-1998 (first entry)  
 XX  
 DE Human BRCA1 omi2 polymorphism #6 cDNA.  
 XX  
 KW BRCA1: omi2; human: breast and ovarian cancer predisposing gene;  
 KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 KW Chromosome 17q; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= "BRCA1 omi2 protein"  
 FT variation 4427  
 FT /\*tag= b  
 FT /note= "This polymorphic variation can be a T or C  
 nucleotide"  
 XX  
 PN US5750400-A.  
 XX  
 PD 12-MAY-1998.  
 XX  
 PE 12-FEB-1997; 97US-0798691.  
 XX  
 PR 12-FEB-1996; 96US-0598591.

PR 12-FEB-1997; 97US-0798691.  
 XX  
 PA (ONCO-) ONCORMED INC.  
 XX  
 PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
 PI Schelter DB, Zeng B;  
 XX  
 DR WP1: 1998-296774/26.  
 XX  
 PT BRCA1 omi2 gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 XX  
 PS Claim 2e; Page -: 54pp; English.  
 CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) omi2 gene in which a polymorphic variation occurs at  
 CC nucleotide 4427. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 omi2 gene represented in AAV46449.  
 XX  
 SO Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;  
 Query Match 100.0%; Score 5709; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCTCGCTGAGACTTCCTGAGACCCGACACAGGCTGTGGGGTTCTCAGATAACTGGGCC 60  
 |||||||  
 DB 1 AGCTCGCTGAGACTTCCTGAGACCCGACACAGGCTGTGGGGTTCTCAGATAACTGGGCC 60  
 QY 61 CCTGGCTCAGAGAGGCTTCCACCCTGTGCTGGGTAAAGTTCAATGGAACAGAAAGAAA 120  
 |||||||  
 DB 61 CCTGGCTCAGAGAGGCTTCCACCCTGTGCTGGGTAAAGTTCAATGGAACAGAAAGAAA 120  
 QY 121 TGGATTTATCTGCTCTTCCGCTTGAAGAGTCAAAATGCTTAATCTTGCAGAAA 180  
 |||||||  
 DB 121 TGGATTTATCTGCTCTTCCGCTTGAAGAGTCAAAATGCTTAATCTTGCAGAAA 180  
 QY 181 TCTTAGAGTGTCCATCTGCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGACC 240  
 |||||||  
 DB 181 TCTTAGAGTGTCCATCTGCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGACC 240  
 QY 241 TCTTAGAGTGTCCATCTGCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGACC 240  
 |||||||  
 DB 241 TCTTAGAGTGTCCATCTGCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGACC 240  
 QY 241 ACATATTTTGGCAATTTTGCATGCTGAACCTTCTACACAGAGAAAGGCTTGCACAGT 300  
 |||||||  
 DB 241 ACATATTTTGGCAATTTTGCATGCTGAACCTTCTACACAGAGAAAGGCTTGCACAGT 300  
 QY 301 GTCCCTTATGTAAGATGATATACCAAAAGAGGCTTCAAGAAAGTGCAGATTGTGTC 360  
 |||||||  
 DB 301 GTCCCTTATGTAAGATGATATACCAAAAGAGGCTTCAAGAAAGTGCAGATTGTGTC 360  
 QY 361 AACCTGTGGAAGAGCTATTTGAAATCATTTTGGCTTTTCAGCTTGACACAGATTGGAGT 420  
 |||||||  
 DB 361 AACCTGTGGAAGAGCTATTTGAAATCATTTTGGCTTTTCAGCTTGACACAGATTGGAGT 420  
 QY 421 ATGCAAAACAGCTATATTTTGGCAAAAAGAAATTAACCTCTGCAACATCTAAAGATG 480  
 |||||||  
 DB 421 ATGCAAAACAGCTATATTTTGGCAAAAAGAAATTAACCTCTGCAACATCTAAAGATG 480  
 QY 481 AAGTTTCTATCATTCGAAGATATGGGCTACAGAAACCGTCAAAAGCTTCTACAGAGTG 540  
 |||||||  
 DB 481 AAGTTTCTATCATTCGAAGATATGGGCTACAGAAACCGTCAAAAGCTTCTACAGAGTG 540  
 QY 541 AACCGAAATATCTTCTTGTGAGAGAAACAGTCTCAGTGTCAACATCTTAACCTTGAA 600  
 |||||||

```
|||||
Db 541 AACCCGAAATCCCTCCCTCCAGAAACCACTGCTCAAGTGTCAACTCTCAACCTTGAA 600
QY CTGTGAACTCTGAGAGCAAGAGGATACAACTCAAAAGAGCTGTCTACATTTG 660
Db 601 CTGTGAACTCTGAGAGCAAGAGGATACAACTCAAAAGAGCTGTCTACATTTG 660
QY AATTGAGTCTGATCTCTGAGAGTACCTTAATAGGCAACTTATTGAGTGGAG 720
Db 661 AATTGAGTCTGATCTCTGAGAGTACCTTAATAGGCAACTTATTGAGTGGAG 720
QY AATTGAGTCTGATCTCTGAGAGTACCTTAATAGGCAACTTATTGAGTGGAG 720
Db 721 ATCAAGATTTGTACAAATACCCCTCAAGAACCAAGGATGAATCAGTTGGATTCTG 780
QY 721 ATCAAGATTTGTACAAATACCCCTCAAGAACCAAGGATGAATCAGTTGGATTCTG 780
Db 781 CAAAAAGGCTGCTGTAATTTCTGAGAGGATTAACAAATACGACATCATAC 840
QY 781 CAAAAAGGCTGCTGTAATTTCTGAGAGGATTAACAAATACGACATCATAC 840
Db 841 CCAATTAATTAATGATTTGAGACCACTGAGAGCTGAGAGGATCCAGAAAGT 900
QY 841 CCAATTAATTAATGATTTGAGACCACTGAGAGCTGAGAGGATCCAGAAAGT 900
Db 841 CCAATTAATTAATGATTTGAGACCACTGAGAGCTGAGAGGATCCAGAAAGT 900
QY 901 ATCAGGCTAGTCTGTTTCAAACTGTCATGTGAGCCATGTGCACAAATCTCATCCA 960
Db 901 ATCAGGCTAGTCTGTTTCAAACTGTCATGTGAGCCATGTGCACAAATCTCATCCA 960
QY 961 GCTCATTTACAGCATGAGACAGCGATTATTACTACATAAAGACAGATTAAGTAAA 1020
Db 961 GCTCATTTACAGCATGAGACAGCGATTATTACTACATAAAGACAGATTAAGTAAA 1020
QY 1021 AGGCTGAATCTGTAATTAAGCAAAACAGCTGCTTGAAGAGGCAATCAATACAT 1080
Db 1021 AGGCTGAATCTGTAATTAAGCAAAACAGCTGCTTGAAGAGGCAATCAATACAT 1080
QY 1081 GGGCTGGAATTAAGCAAAACATGTAATGATAGGGGACTCCACAGACAGAAAAAGGTAG 1140
Db 1081 GGGCTGGAATTAAGCAAAACATGTAATGATAGGGGACTCCACAGACAGAAAAAGGTAG 1140
QY 1141 ATCTGAATGCTGATCCCTGCTGAGAGAAAGATGAATTAAGCAAAAGTGCATGCT 1200
Db 1141 ATCTGAATGCTGATCCCTGCTGAGAGAAAGATGAATTAAGCAAAAGTGCATGCT 1200
QY 1201 CAGAGATCTCTAGAGATCTGAGATGTCCTTGGATTAACATAATAGCAGATTGAGA 1260
Db 1201 CAGAGATCTCTAGAGATCTGAGATGTCCTTGGATTAACATAATAGCAGATTGAGA 1260
QY 1261 AAGTTAATGATGCTTTTCCAGAAAGTATCACTGTTAGTTCATGATGACTCACATGATG 1320
Db 1261 AAGTTAATGATGCTTTTCCAGAAAGTATCACTGTTAGTTCATGATGACTCACATGATG 1320
QY 1321 GGGAGTCTGAATCAAAATGCCAAAGTATGATGATGAGCTTCTAAATGAGGTAGATG 1380
Db 1321 GGGAGTCTGAATCAAAATGCCAAAGTATGATGATGAGCTTCTAAATGAGGTAGATG 1380
QY 1381 AATATCTGCTCTCTGAGAGAAATAGACTTACGAGCAGTGAATCCATGAGGCTTTAA 1440
Db 1381 AATATCTGCTCTCTGAGAGAAATAGACTTACGAGCAGTGAATCCATGAGGCTTTAA 1440
QY 1441 TATGTAATGATGAAAGTCTCACTCAATCAGTAGAGATTAATATTGAAGACAAATAT 1500
Db 1441 TATGTAATGATGAAAGTCTCACTCAATCAGTAGAGATTAATATTGAAGACAAATAT 1500
QY 1501 TTGGGAAACCTTATCGGAGAGCAAGGCTCCCAATTAAGCATGTAAGTGAATATC 1560
Db 1501 TTGGGAAACCTTATCGGAGAGCAAGGCTCCCAATTAAGCATGTAAGTGAATATC 1560
QY 1561 TAAATTAAGAGCATTTGTTACTGAGCAGACAGATTAATCAAGAGCGTCCCTCACAATA 1620
Db 1561 TAAATTAAGAGCATTTGTTACTGAGCAGACAGATTAATCAAGAGCGTCCCTCACAATA 1620
QY 1621 AATTTAAAGCTAAAGAGACCTACATCAGGCTTCACTCTGAGGATTTTATCAAGAAAG 1680
Db 1621 AATTTAAAGCTAAAGAGACCTACATCAGGCTTCACTCTGAGGATTTTATCAAGAAAG 1680

|||||
Db 1621 AATTTAAAGCTAAAGAGACCTACATCAGGCTTCACTCTGAGGATTTTATCAAGAAAG 1680
QY 1681 CAGATTTGGCAGTCTCAAAAGACCTCTGAAATGATTAATCAGGAGTACCAAAAGCGAGC 1740
Db 1681 CAGATTTGGCAGTCTCAAAAGACCTCTGAAATGATTAATCAGGAGTACCAAAAGCGAGC 1740
QY 1741 AGAATGCTCAAGTGAATTAATTAAGTGTCTGATGAGATTAATCAAAAGTGAAT 1800
Db 1741 AGAATGCTCAAGTGAATTAATTAAGTGTCTGATGAGATTAATCAAAAGTGAAT 1800
QY 1801 CTATTCAGATGAGAAAAATCTTAACCAATAGATAGATCAGTCAAAAAAGATCTCTTCA 1860
Db 1801 CTATTCAGATGAGAAAAATCTTAACCAATAGATAGATCAGTCAAAAAAGATCTCTTCA 1860
QY 1861 AAACGAAGCTGAACTTAAGAGCAGTAAACCAATATGGAATGGAATTAATATTC 1920
Db 1861 AAACGAAGCTGAACTTAAGAGCAGTAAACCAATATGGAATGGAATTAATATTC 1920
QY 1921 ACAATTCAAAAAGCCTTAAGAAAGATAGGCTGAGAGAGGATCTTACAGGATATTC 1980
Db 1921 ACAATTCAAAAAGCCTTAAGAAAGATAGGCTGAGAGAGGATCTTACAGGATATTC 1980
QY 1981 ATGCGCTTGAATGATAGTACGATGAATCTAAGCCCACTAATTGTTACTGATTCGAAA 2040
Db 1981 ATGCGCTTGAATGATAGTACGATGAATCTAAGCCCACTAATTGTTACTGATTCGAAA 2040
QY 2041 TTGATGATTTGTTCTAGAGAGTGAAGATTAAGAAAAAGTACAAACCAATGGCAGTCA 2100
Db 2041 TTGATGATTTGTTCTAGAGAGTGAAGATTAAGAAAAAGTACAAACCAATGGCAGTCA 2100
QY 2101 GGCACAGCAGAAAACTTCAACATCTGAGAAAGGTAAAGAACCTGCACTGAGACCAAGAGA 2160
Db 2101 GGCACAGCAGAAAACTTCAACATCTGAGAAAGGTAAAGAACCTGCACTGAGACCAAGAGA 2160
QY 2161 GTAACAAAGCCAAATGAACAGACAGTAAACATGACAGGATCTTCCACAGAGTGA 2220
Db 2161 GTAACAAAGCCAAATGAACAGACAGTAAACATGACAGGATCTTCCACAGAGTGA 2220
QY 2221 AGTTAAATGCTGAGAGACCTGCTTCTTCTAAGTCTCAATATACAGGAACTTAAAGAT 2280
Db 2221 AGTTAAATGCTGAGAGACCTGCTTCTTCTAAGTCTCAATATACAGGAACTTAAAGAT 2280
QY 2281 TTGTCAATCTAGGCTTCCAGAGAGAAAAAGAGAAACCTAGAAACACTTAAAGTGT 2340
Db 2281 TTGTCAATCTAGGCTTCCAGAGAGAAAAAGAGAAACCTAGAAACACTTAAAGTGT 2340
QY 2341 CTAAATATGCTGAGAGACCTCAAGATCTCATGTTAAGTGAAGAGGTTTGGCAACTG 2400
Db 2341 CTAAATATGCTGAGAGACCTCAAGATCTCATGTTAAGTGAAGAGGTTTGGCAACTG 2400
QY 2401 AAAGATCTGAGAGAGTGAAGCTATTTCAATGTTGATGCTGATGATTAAGCTGAG 2460
Db 2401 AAAGATCTGAGAGAGTGAAGCTATTTCAATGTTGATGCTGATGATTAAGCTGAG 2460
QY 2461 AAAGATCTGAGAGAGTGAAGCTATTTCAATGTTGATGCTGATGATTAAGCTGAG 2520
Db 2461 AAAGATCTGAGAGAGTGAAGCTATTTCAATGTTGATGCTGATGATTAAGCTGAG 2520
QY 2521 GTGTGATGCTGAGTGTGAGCATTTTAAAGCCCAAGGAGCTAATTCATGTTGTTGCAAG 2580
Db 2521 GTGTGATGCTGAGTGTGAGCATTTTAAAGCCCAAGGAGCTAATTCATGTTGTTGCAAG 2580
QY 2581 ATATTAAGATGACAGAGAGGCTTTAAGTATTCATGTTGAGCATGATTAAGCATGATC 2640
Db 2581 ATATTAAGATGACAGAGAGGCTTTAAGTATTCATGTTGAGCATGATTAAGCATGATC 2640
QY 2641 GGGAAACAGCATGAGAAATGAGAAAGTGAACCTTGAATGCTCAGATTTGAGCAATACAT 2700
Db 2641 GGGAAACAGCATGAGAAATGAGAAAGTGAACCTTGAATGCTCAGATTTGAGCAATACAT 2700
QY 2701 TCAAGGTTTCAAGGCGGAGCTGATGCTGTTTCAAAATCCAGAAATGACAGAAAGG 2760
Db 2701 TCAAGGTTTCAAGGCGGAGCTGATGCTGTTTCAAAATCCAGAAATGACAGAAAGG 2760
```

QY	2761	AATGTGCAACATTTCTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAAGTCACTT	2820
Db	2761	AATGTGCAACATTTCTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAAGTCACTT	2820
QY	2821	TTGCAATGTGAACAAAAGAGAAATTCAAAGGAAAGAATGAGTCTAATATCAAGCCTGTAC	2880
Db	2821	TTGCAATGTGAACAAAAGAGAAATTCAAAGGAAAGAATGAGTCTAATATCAAGCCTGTAC	2880
QY	2881	AGACAGTTAATATCACTCGAGGCTTTCTGTGTGTGGTGCAGAAAGATTAAGCAGTTGTATA	2940
Db	2881	AGACAGTTAATATCACTCGAGGCTTTCTGTGTGTGGTGCAGAAAGATTAAGCAGTTGTATA	2940
QY	2941	ATGCCAAATGTAGTATCAAGAGGCTAGGTTTTGTCTATCATCTCAGTTCAAGGCA	3000
Db	2941	ATGCCAAATGTAGTATCAAGAGGCTAGGTTTTGTCTATCATCTCAGTTCAAGGCA	3000
QY	3001	ACGAACATGAGACTCTTACTCTCCAAATTAACATGAGCTTTTCCAAAAACCATATGCTATAC	3060
Db	3001	ACGAACATGAGACTCTTACTCTCCAAATTAACATGAGCTTTTCCAAAAACCATATGCTATAC	3060
QY	3061	CACCACTTTTCCATCAAGTCAATTTGTTAAACATAATGTAAAGAAATCTGCTAGAG	3120
Db	3061	CACCACTTTTCCATCAAGTCAATTTGTTAAACATAATGTAAAGAAATCTGCTAGAG	3120
QY	3121	AAAATTTGAGGAACATTCATCATGTCCCTGAAAGAAATGGAATGAGAAACATTCOA	3180
Db	3121	AAAATTTGAGGAACATTCATCATGTCCCTGAAAGAAATGGAATGAGAAACATTCOA	3180
QY	3181	GTACAGTAGACCAATTTAGCCGTAAATTAACATTAAGAAATGTTTTTAAGAACGCACT	3240
Db	3181	GTACAGTAGACCAATTTAGCCGTAAATTAACATTAAGAAATGTTTTTAAGAACGCACT	3240
QY	3241	CAAGCAATATTATGAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA	3300
Db	3241	CAAGCAATATTATGAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA	3300
QY	3301	TAGGTTCCAGTAGTAAACAAATTCOAAGAGAACTAGTAGAAACAGAGGCGCAAAATGCA	3360
Db	3301	TAGGTTCCAGTAGTAAACAAATTCOAAGAGAACTAGTAGAAACAGAGGCGCAAAATGCA	3360
QY	3361	ATGCTATCTTAGATTAGGGGTTTTTGGCAACCTGAGGCTATATAACAAAGTCTTCTGTGAA	3420
Db	3361	ATGCTATCTTAGATTAGGGGTTTTTGGCAACCTGAGGCTATATAACAAAGTCTTCTGTGAA	3420
QY	3421	GTAATTTGTAACATCTCTGAAATTAACAAAGCAAGATTAAGAAAGTAAGTACTAGACTGTGA	3480
Db	3421	GTAATTTGTAACATCTCTGAAATTAACAAAGCAAGATTAAGAAAGTAAGTACTAGACTGTGA	3480
QY	3481	ATACAGATTTCTCTCATATCTGATTTGAGATTACTTGAACAGAGCTTTGGGAAGTATC	3540
Db	3481	ATACAGATTTCTCTCATATCTGATTTGAGATTACTTGAACAGAGCTTTGGGAAGTATC	3540
QY	3541	ATGCATCTCAGGTTTTGTTCTGAGACACTGATGACCTGTGTAGATGATGCTGAATTAAGG	3600
Db	3541	ATGCATCTCAGGTTTTGTTCTGAGACACTGATGACCTGTGTAGATGATGCTGAATTAAGG	3600
QY	3601	AAGATACTAGTTTGTGCAAAATGCAATTAAGGAAGTGTGCTGTTTTGCAAAAGCG	3660
Db	3601	AAGATACTAGTTTGTGCAAAATGCAATTAAGGAAGTGTGCTGTTTTGCAAAAGCG	3660
QY	3661	TCGAAAGAGAGAGCTTACAGAGAGTCCCTACCCCTTACCCCATACACTTTGGCTCAGG	3720
Db	3661	TCGAAAGAGAGAGCTTACAGAGAGTCCCTACCCCTTACCCCATACACTTTGGCTCAGG	3720
QY	3721	GTTACCGAAGAGGGGCCAAGAAATTTAGAGTCTCTCAGAAAGAACTTATCTAGTAGAGATG	3780
Db	3721	GTTACCGAAGAGGGGCCAAGAAATTTAGAGTCTCTCAGAAAGAACTTATCTAGTAGAGATG	3780
QY	3781	AAGAGCTTCCCTGCTTCCAAACACTTGTTATTGTAAGGTAACAAATATACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTTCCAAACACTTGTTATTGTAAGGTAACAAATATACCTTCTCAGT	3840

QY	3841	CTACTAGCATATGACACCGTTCGTACCAGCTGTCTCTCTAATAACAACAGAGAAATTTAT	3900
Db	3841	CTACTAGGCATATGACACCGTTCGTACCAGGTGTCTCTCTAATAACAACAGAGAAATTTAT	3900
QY	3901	TATCATTTGAAGAATATGCTTTAAATGACTCAGTAACACAGTATATATTTGGCAAGCATCTC	3960
Db	3901	TATCATTTGAAGAATATGCTTTAAATGACTCAGTAACACAGTATATATTTGGCAAGCATCTC	3960
QY	3961	AGGAACATCACCTTATAGTAGAGAAACAAATGTCTGCTAGCTTTCTTCCACAGTGA	4020
Db	3961	AGGAACATCACCTTATAGTAGAGAAACAAATGTCTGCTAGCTTTCTTCCACAGTGA	4020
QY	4021	GTGAATTTGGAGACCTTGACTGCAATATACAAACACCCAGATCTTTCTTGATTTGGTCTT	4080
Db	4021	GTGAATTTGGAGACCTTGACTGCAATATACAAACACCCAGATCTTTCTTGATTTGGTCTT	4080
QY	4081	CCAAACAATATGAGCATAGTCTGAGTAACCCAGGGAGTGGTCTGAGTACAGGAATTTGG	4140
Db	4081	CCAAACAATATGAGCATAGTCTGAGTAACCCAGGGAGTGGTCTGAGTACAGGAATTTGG	4140
QY	4141	TTTTGATGATGTAAGAAAGAGGAACGGGCTTTGGAAGAAATTAATCAAGAACGCAAGCA	4200
Db	4141	TTTTGATGATGTAAGAAAGAGGAACGGGCTTTGGAAGAAATTAATCAAGAACGCAAGCA	4200
QY	4201	TGGATTTCAAACCTTAGTGAAGACAGCATCTGGGTGTGAGAGTGAACAACGCTCTCTAAG	4260
Db	4201	TGGATTTCAAACCTTAGTGAAGACAGCATCTGGGTGTGAGAGTGAACAACGCTCTCTAAG	4260
QY	4261	ACTGTCAGGGGTATTCCTCTCAGATGTGACATTTTAACCATCTACAGAGGGATACCATGC	4320
Db	4261	ACTGTCAGGGGTATTCCTCTCAGATGTGACATTTTAACCATCTACAGAGGGATACCATGC	4320
QY	4321	AACATTAACCTGATTAAGCTCCAGCAGGAATATGGCTGAACCTAGAGCTGTGTTAACAAGC	4380
Db	4321	AACATTAACCTGATTAAGCTCCAGCAGGAATATGGCTGAACCTAGAGCTGTGTTAACAAGC	4380
QY	4381	ATGGAGCCAGGCTTCTTAACAGACTACCTCCATCATTAAGTACTCTTCTGCCCCCTGAGG	4440
Db	4381	ATGGAGCCAGGCTTCTTAACAGACTACCTCCATCATTAAGTACTCTTCTGCCCCCTGAGG	4440
QY	4441	ACCTCGGAATTCACAAACAAAGGCACATGAAAAAGCAAGTATTAACCTTCACAGAAAGTA	4500
Db	4441	ACCTCGGAATTCACAAACAAAGGCACATGAAAAAGCAAGTATTAACCTTCACAGAAAGTA	4500
QY	4501	GTGAATACCCTATTAAGCCAGAAATCCAGAAAGCCTTTCGCTGACAAAGTTTGAAGTGTCTG	4560
Db	4501	GTGAATACCCTATTAAGCCAGAAATCCAGAAAGCCTTTCGCTGACAAAGTTTGAAGTGTCTG	4560
QY	4561	CAGATAGTTCTPAACAGTAAAAATAAAGAACCAAGAGTGGAAAGGTCAATCCCTCTTAAT	4620
Db	4561	CAGATAGTTCTPAACAGTAAAAATAAAGAACCAAGAGTGGAAAGGTCAATCCCTCTTAAT	4620
QY	4621	GCCCATCAATTATGATATAGTGTGATCAAGCAGTGTGCTGGGAGTCTTCAGAAATGAA	4680
Db	4621	GCCCATCAATTATGATATAGTGTGATCAAGCAGTGTGCTGGGAGTCTTCAGAAATGAA	4680
QY	4681	ACTAACCCATCTCAAGAGAGAGCTCAATTAAAGTGTGTTGAATGTGAGAGACCAACAGCTGGAAG	4740
Db	4681	ACTAACCCATCTCAAGAGAGAGCTCAATTAAAGTGTGTTGAATGTGAGAGACCAACAGCTGGAAG	4740
QY	4741	AGTCTGGGCCACACGATTTGACGGAACATCTTACTTGCCAAAGCAGATCTGAGGGAA	4800
Db	4741	AGTCTGGGCCACACGATTTGACGGAACATCTTACTTGCCAAAGCAGATCTGAGGGAA	4800
QY	4801	CCCCCTTACCTTGAATCTGGAAATCTGGAATCAAGCCTCTCTCTGATGACCTTGATCTCTTCTG	4860
Db	4801	CCCCCTTACCTTGAATCTGGAAATCTGGAATCAAGCCTCTCTCTGATGACCTTGATCTCTTCTG	4860
QY	4861	AAGACAGAGCCCCAGAGTCAAGCTGTGTGTGGCAACATACATCTTCAACCTCTGCAATTA	4920
Db	4861	AAGACAGAGCCCCAGAGTCAAGCTGTGTGTGGCAACATACATCTTCAACCTCTGCAATTA	4920
QY	4921	AAGTTTCCCAATTAAGTTGCAAAATCTGCCAGATCCAGCTGCTCATACTACTG	4980

QY	2521	GTGTCAGCATGTCGACGACATTTTGAAAAACCCCAAGGACATTAATTCATGTTGTTCCAAAG	2580
Db	2521	GTGTGAGTCAGATGTGACACATTTTAAAAACCCCAAGGACATTAATTCATGTTGTTCCAAAG	2580
QY	2581	ATATATAGAAAATGACACAGAAAGGCTTTTAAGTTCATTTGGGACATGAAGTTTAACCAAGTC	2640
Db	2581	ATATATAGAAAATGACACAGAAAGGCTTTTAAGTTCATTTGGGACATGAAGTTTAACCAAGTC	2640
QY	2641	GGGAACACAGCATGGAATGGAGAAAGTGAACCTTGATGCTCATTTTGCAGAAATCAT	2700
Db	2641	GGGAACACAGCATGGAATGGAGAAAGTGAACCTTGATGCTCATTTTGCAGAAATCAT	2700
QY	2701	TCAAGGTTTCAAAAGCGCCAGTCATTTGCTCTGTTTTCAAATCCAGAAATCCAGAGAGG	2760
Db	2701	TCAAGGTTTCAAAAGCGCCAGTCATTTGCTCTGTTTTCAAATCCAGAAATCCAGAGAGG	2760
QY	2761	AATGTCGAACATTCCTGCTCCCACTGTGGGCTCTTAAGAAACCAAGTCCAAAGTCACCTT	2820
Db	2761	AATGTCGAACATTCCTGCTCCCACTGTGGGCTCTTAAGAAACCAAGTCCAAAGTCACCTT	2820
QY	2821	TTGAATGTGAACAAAAGGAGAAATAATCAGGAAGAATGACTTAATATCAAGCCTGTAC	2880
Db	2821	TTGAATGTGAACAAAAGGAGAAATAATCAGGAAGAATGACTTAATATCAAGCCTGTAC	2880
QY	2881	AGACAGTTAATATATACATGACGAGGCTTCTGTGTGTTGGTCAGAAAGATTAAGCCAGTTGATA	2940
Db	2881	AGACAGTTAATATATACATGACGAGGCTTCTGTGTGTTGGTCAGAAAGATTAAGCCAGTTGATA	2940
QY	2941	ATGCGCAATGAGTATGATCAAAAGGAGGCTGTAGGTTTGCTATCATCTCAGTTCAGAGGCA	3000
Db	2941	ATGCGCAATGAGTATGATCAAAAGGAGGCTGTAGGTTTGCTATCATCTCAGTTCAGAGGCA	3000
QY	3001	ACGAAGCTGCACTCATTACTCCAAATTAACATGAGACTTTTACAAACCCCATATGCTATAC	3060
Db	3001	ACGAAGCTGCACTCATTACTCCAAATTAACATGAGACTTTTACAAACCCCATATGCTATAC	3060
QY	3061	CACCACCTTTTCCCATCAAGTCATTTGTTTAAACTAATGTAGAAAAATCTGCTAGAGG	3120
Db	3061	CACCACCTTTTCCCATCAAGTCATTTGTTTAAACTAATGTAGAAAAATCTGCTAGAGG	3120
QY	3121	AAAACCTTGGAGAAATTCATGTCACCTGTAAGACGAATGGGAAATGAGAACCTTCCAA	3180
Db	3121	AAAACCTTGGAGAAATTCATGTCACCTGTAAGACGAATGGGAAATGAGAACCTTCCAA	3180
QY	3181	GTACAGTGAGACACAAATTAAGCCGTAAATTAACATTTAGAGAAAATGTTTAAAGAACCCAGCT	3240
Db	3181	GTACAGTGAGACACAAATTAAGCCGTAAATTAACATTTAGAGAAAATGTTTAAAGAACCCAGCT	3240
QY	3241	CAAGCAATTAATTAAGAGTAGTGTCCAGTACTAATGAAGTAGGCGCTCCAGTAATTAATGAAA	3300
Db	3241	CAAGCAATTAATTAAGAGTAGTGTCCAGTACTAATGAAGTAGGCGCTCCAGTAATTAATGAAA	3300
QY	3301	TAGGTTCCAGTGAGAAACAAATTCACACAGAACCTAGGTAGAAACAGAGGCGCAAAATTTGA	3360
Db	3301	TAGGTTCCAGTGAGAAACAAATTCACACAGAACCTAGGTAGAAACAGAGGCGCAAAATTTGA	3360
QY	3361	ATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTCTATAAACCAAGTCTTCTTGAAA	3420
Db	3361	ATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTCTATAAACCAAGTCTTCTTGAAA	3420
QY	3421	GTAATTTGTAAGCATCTGGAATAAAAAAGCAAGATATGAAGAAAGTAGTTCAGACTGTTA	3480
Db	3421	GTAATTTGTAAGCATCTGGAATAAAAAAGCAAGATATGAAGAAAGTAGTTCAGACTGTTA	3480
QY	3481	ATTACAGATTTCTCTCCATATCTGATTTTCACATTAATCTAGAACACGCTATGGGAAGTAGTC	3540
Db	3481	ATTACAGATTTCTCTCCATATCTGATTTTCACATTAATCTAGAACACGCTATGGGAAGTAGTC	3540
QY	3541	ATGCATCTCAAGGTTGTTGTCGAGACACCTGATGACCTGTTAGATGATGATGGTGAATTAAGG	3600
Db	3541	ATGCATCTCAAGGTTGTTGTCGAGACACCTGATGACCTGTTAGATGATGATGGTGAATTAAGG	3600
QY	3601	AAGATTAAGTTTGTCTGAAAATGACATTAAGAAAGTCTGTGCTGTTTAAAGCAAAAAGCG	3660

Dp	3601	AAGATAGCTAGTTTGGCTGAAAAATGACATTAAGGAAAGTTCTGCTTTTGTAGCAAAACCG	3660
Qy	3661	TCCAAAGAGAGAGCTTGTAGCAGAGATCTTAGCCCTTTACCCATTCACATTTGGCTAGG	3720
Dp	3661	TCCAAAGAGAGAGCTTGTAGCAGAGATCTTAGCCCTTTACCCATTCACATTTGGCTAGG	3720
Qy	3721	GTTACCGAAGAGGGGCCAAAGAAATTAGAGTCTCGAAGAGAACCTTATCTAGTGAGATG	3780
Dp	3721	GTTACCGAAGAGGGGCCAAAGAAATTAGAGTCTCGAAGAGAACCTTATCTAGTGAGATG	3780
Qy	3781	AAGAGCTTCCCTGCTTCCACACTTGTATTGTGAAGTAACATATPACCTTCTCAGT	3840
Dp	3781	AAGAGCTTCCCTGCTTCCACACTTGTATTGTGAAGTAACATATPACCTTCTCAGT	3840
Qy	3841	CTACTAGGCAATAGCACCGCTGTCTACCGAGTGTCTGTAAAGAACACAGAGAAATTTAT	3900
Dp	3841	CTACTAGGCAATAGCACCGCTGTCTACCGAGTGTCTGTAAAGAACACAGAGAAATTTAT	3900
Qy	3901	TATCATTTGAAGAAATAGCTTAAATGACTCCAGTAACAGAGTAATATTGGCAAGCATCTC	3960
Dp	3901	TATCATTTGAAGAAATAGCTTAAATGACTCCAGTAACAGAGTAATATTGGCAAGCATCTC	3960
Qy	3961	AGGAACATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTTGTTCTTACAGTGCA	4020
Dp	3961	AGGAACATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTTGTTCTTACAGTGCA	4020
Qy	4021	GTCATTTGGAAGACTTGACTGCAATACAAACACCACAGATCCCTTGTGATGGTCTT	4080
Dp	4021	GTCATTTGGAAGACTTGACTGCAATACAAACACCACAGATCCCTTGTGATGGTCTT	4080
Qy	4081	CCAAACAAATGAGGCACTGACTGTAAGCCACAGGGAGTTGGTCTGAGTAGACAAAGAAATGG	4140
Dp	4081	CCAAACAAATGAGGCACTGACTGTAAGCCACAGGGAGTTGGTCTGAGTAGACAAAGAAATGG	4140
Qy	4141	TTTCAGATGATGAGAAAGAGAGAACGGGCTTGGAAAGAAATATATCAGAGAGCAACCA	4200
Dp	4141	TTTCAGATGATGAGAAAGAGAGAACGGGCTTGGAAAGAAATATATCAGAGAGCAACCA	4200
Qy	4201	TGGATTTCAACTTAGTAGAGAGCATCGGTGGTGTGAGAGTGAACCAAGCCCTCTGTGAG	4260
Dp	4201	TGGATTTCAACTTAGTAGAGAGCATCGGTGGTGTGAGAGTGAACCAAGCCCTCTGTGAG	4260
Qy	4261	ACTGCTCAGGGCTATCTCTCAGATGACATTTTAAACACATCAGAGAGGATACCATG	4320
Dp	4261	ACTGCTCAGGGCTATCTCTCAGATGACATTTTAAACACATCAGAGAGGATACCATG	4320
Qy	4321	AACATTAACCTGATTAAGCTCCAGCAGGAATGGCTGAACTAGAAAGCTGTGTAGAACGC	4380
Dp	4321	AACATTAACCTGATTAAGCTCCAGCAGGAATGGCTGAACTAGAAAGCTGTGTAGAACGC	4380
Qy	4381	ATGGAGGCCACCTTCTTAAGCAAGCTTAACCTTCCATCATTAAGTACCTTCTGCTTGAG	4440
Dp	4381	ATGGAGGCCACCTTCTTAAGCAAGCTTAACCTTCCATCATTAAGTACCTTCTGCTTGAG	4440
Qy	4441	ACCTGCGAAATTCAGAAACAAAGACATAGAAAAAGCAGTATTAACTTCACAGAAAGCTA	4500
Dp	4441	ACCTGCGAAATTCAGAAACAAAGACATAGAAAAAGCAGTATTAACTTCACAGAAAGCTA	4500
Qy	4501	GTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCTTTTCTGCTGACAAAGTTTGAGGTCTG	4560
Dp	4501	GTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCTTTTCTGCTGACAAAGTTTGAGGTCTG	4560
Qy	4561	CAGATAGCTTACACAGTAAATAATTAAGAACCAAGAGTGGAAAGGTCAATCCCTTCTTAAT	4620
Dp	4561	CAGATAGCTTACACAGTAAATAATTAAGAACCAAGAGTGGAAAGGTCAATCCCTTCTTAAT	4620
Qy	4621	GCCCATCTTATAGATATAGGTGTATACATGACACAGTGTCTTGAGAGTCTTACAGATAGAA	4680
Dp	4621	GCCCATCTTATAGATATAGGTGTATACATGACACAGTGTCTTGAGAGTCTTACAGATAGAA	4680
Qy	4681	ACTACCCATCTCAAGAGAGCTCATTTAAGCTTGTGATGTGAGAGAGCAACAGCTGGAAG	4740
Dp	4681	ACTACCCATCTCAAGAGAGCTCATTTAAGCTTGTGATGTGAGAGAGCAACAGCTGGAAG	4740

```

Db 4681 ACTACCATCTCAAGAGAGACCTATTAGGTTGTGATGTGAGAGCAACAGCTGGAG 4740
QY 4741 AGTCTGGCCACACGATTTTACGGAACATCTTACTTGTCCAGAGCATGTAGAGGAA 4800
Db 4741 AGTCTGGCCACACGATTTTACGGAACATCTTACTTGTCCAGAGCATGTAGAGGAA 4800
QY 4801 CCCCTTACCTGGAATGTGAATCAGCCCTTCTCTGTATGACCCCTGAATGTATCTTCTG 4860
Db 4801 CCCCTTACCTGGAATGTGAATCAGCCCTTCTCTGTATGACCCCTGAATGTATCTTCTG 4860
QY 4861 AAGACAGAGCCCAAGATGACGTCGTCGTGTGGCAACATACCATCTTCAACCTGTGATTTGA 4920
Db 4861 AAGACAGAGCCCAAGATGACGTCGTCGTGTGGCAACATACCATCTTCAACCTGTGATTTGA 4920
QY 4921 AAGTTCGCCAATGAAGTTGCGAATCTGCCAGATCCAGCTGCTCATACTACTG 4980
Db 4921 AAGTTCGCCAATGAAGTTGCGAATCTGCCAGATCCAGCTGCTCATACTACTG 4980
QY 4981 ATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGACAGGAGAAAGCCAGAAATGACAG 5040
Db 4981 ATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGACAGGAGAAAGCCAGAAATGACAG 5040
QY 5041 CTTCAACAGAAAGGTCACAAAGAAATGTCATGGTGTGTGCTGCTTACCCAGAG 5100
Db 5041 CTTCAACAGAAAGGTCACAAAGAAATGTCATGGTGTGTGCTGCTTACCCAGAG 5100
QY 5101 AATTATGCTGCTGTAAGTTGCGAAGAACCCACATCATTAACTAATTA 5160
Db 5101 AATTATGCTGCTGTAAGTTGCGAAGAACCCACATCATTAACTAATTA 5160
QY 5161 CTGAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGAGAC 5220
Db 5161 CTGAGAGACTACTCATGTTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGAGAC 5220
QY 5221 TGAATATTTTCTAGGAATTTGGGAGGAAATGGGAGTACTATTTCTGGTGACC 5280
Db 5221 TGAATATTTTCTAGGAATTTGGGAGGAAATGGGAGTACTATTTCTGGTGACC 5280
QY 5281 AGCTATTTAAAGAAAGAAATGCTGAATGAGCATATTTTGAAGTCAGAGAGATGTG 5340
Db 5281 AGCTATTTAAAGAAAGAAATGCTGAATGAGCATATTTTGAAGTCAGAGAGATGTG 5340
QY 5341 TCAATGGAAGAAACCAAGGTCACAAAGCGAGAGAGATCCAGAGAGAAAGATCT 5400
Db 5341 TCAATGGAAGAAACCAAGGTCACAAAGCGAGAGAGATCCAGAGAGAAAGATCT 5400
QY 5401 TCAGGGGGCTAGAAATCTGTGCTATGGGCCCTTCAACAACATGCCACAGATCACTGG 5460
Db 5401 TCAGGGGGCTAGAAATCTGTGCTATGGGCCCTTCAACAACATGCCACAGATCACTGG 5460
QY 5461 AATGATGTTACAGCTGTGTGTGCTGTGTGTAAGAGAGCTTTCATCTTACCTTGG 5520
Db 5461 AATGATGTTACAGCTGTGTGTGCTGTGTGTAAGAGAGCTTTCATCTTACCTTGG 5520
QY 5521 GCACAGGTTCACACCAATTTGTGTGTGACAGCAGATGCTTGGACAGAGACATGGCT 5580
Db 5521 GCACAGGTTCACACCAATTTGTGTGTGACAGCAGATGCTTGGACAGAGACATGGCT 5580
QY 5581 TCCATGCAATTTGGGAGATGTGAGGACCTGTGTGACCCAGAGTGGTGTGGACA 5640
Db 5581 TCCATGCAATTTGGGAGATGTGAGGACCTGTGTGACCCAGAGTGGTGTGGACA 5640
QY 5641 GTGTAGCACTTACAGTGTCCAGAGACTGACACTACCTGATACCCAGATCCCCACA 5700
Db 5641 GTGTAGCACTTACAGTGTCCAGAGACTGACACTACCTGATACCCAGATCCCCACA 5700
QY 5701 GCCACTACTGA 5711
Db 5701 GCCACTACTGA 5711

```

RESULT 7  
AAV46458

```

ID AAV46458 standard; cDNA: 5711 BP.
XX
AC AAV46458;
XX
DT 18-NOV-1998 (first entry)
XX
DE Human BRCA1 omi2 polymorphism #1 cDNA.
XX
KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;
KM polymorphism; susceptibility; anti-oncogene; tumour suppressor;
XX chromosome 17q; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT FT /*tag= a
FT FT /product= "BRCA1 omi2 protein"
FT FT variation 2201
FT FT /*tag= b
FT FT /note= "this polymorphic variation can be a C or T
nucleotide"
XX
XX US5750400-A.
XX
XX 12-MAY-1998.
XX
XX 12-FEB-1997; 9705-0798691.
XX
XX 12-FEB-1996; 9605-0598591.
XX
XX 12-FEB-1997; 9705-0798691.
XX
XX (ONCO-) ONCORMED INC.
XX
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
PI Scheller DB, Zeng B;
XX
XX WPI: 1998-296774/26.
XX
PT BRCA1 omi2 gene coding sequences - useful for distinguishing between
PT polymorphisms and mutation(s) in the screening for disposition to
PT breast or ovarian cancer
XX
XX PS Claim 2e; Page -: 54pp; English.
XX
CC This sequence encodes a human BRCA1 (breast and ovarian cancer
CC predisposing gene) omi2 gene in which a polymorphic variation occurs at
CC nucleotide 2201. This sequence and other polymorphic variations of this
CC sequence are useful for the identification of an individual who may or
CC may not have an increased susceptibility to breast or ovarian cancer.
CC The sequences used identify gene changes which are due to polymorphisms,
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC suppressor) which is involved in genetic inheritance of cancers,
CC especially breast and ovarian cancer. It is found at human chromosome
CC 17q which is known to be linked to cancer susceptibility, especially
CC breast cancer. Cells containing a mutation in this gene lose the
CC wild-type function of BRCA1 and are more susceptible to cancers.
CC NOTE: This sequence does not appear in the specification but has been
CC created from the wild type BRCA1 omi2 gene represented in AAV46449.
XX
XX
SQ Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;
XX
Query Match 100.0%; Score 5709; DB 19; Length 5711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGCTGCTGAGACTTCTGAGCCCGCCAGAGCTGTGGGTTTCTGAGATACTGGCC 60
Db 1 AGCTGCTGAGACTTCTGAGCCCGCCAGAGCTGTGGGTTTCTGAGATACTGGCC 60
QY 61 CCTGGCCTCAGAGAGGCTTACCCCTTCTGTGCTGTGGGTAAGTTCAATGGAAGAAAGAA 120
Db 61 CCTGGCCTCAGAGAGGCTTACCCCTTCTGTGCTGTGGGTAAGTTCAATGGAAGAAAGAA 120

```

121 TGAATTTATCTGCTTCGCGCTTGAAGAAGTACAAATGTCATTATGCTATGAGAAA 180  
121 TGAATTTATCTGCTTCGCGCTTGAAGAAGTACAAATGTCATTATGCTATGAGAAA 180  
181 TCTTAGAGTGTCCATCTGTGTGAGTTGATCAAGAACTGTGTCCCAAAAGTGTAGC 240  
181 TCTTAGAGTGTCCATCTGTGTGAGTTGATCAAGAACTGTGTCCCAAAAGTGTAGC 240  
241 ACATATTTTGAATTTTGCATGCTGAACCTTCAACCCAGAGAAAGGCGCTCACAGT 300  
241 ACATATTTTGAATTTTGCATGCTGAACCTTCAACCCAGAGAAAGGCGCTCACAGT 300  
301 GTCCCTTATGTAGAATGATATAACCAAAAGAGCCTTACAGAAAGTACAGATTTATGTC 360  
301 GTCCCTTATGTAGAATGATATAACCAAAAGAGCCTTACAGAAAGTACAGATTTATGTC 360  
361 AACTTGTGAAGAGCTTATGAAAATCATTTGTGCTTTTACGCTTGTACACAGGTTGAGT 420  
361 AACTTGTGAAGAGCTTATGAAAATCATTTGTGCTTTTACGCTTGTACACAGGTTGAGT 420  
421 ATGCAACAGCTATATTTTGCATGCTGAACCTTCAACCCAGAGAAAGGCGCTCACAGT 480  
421 ATGCAACAGCTATATTTTGCATGCTGAACCTTCAACCCAGAGAAAGGCGCTCACAGT 480  
481 AAGTTTCTATCAGCAAGTATGAGCTACAGAAACCGTGCACAAAGACTTCTACAGAGTG 540  
481 AAGTTTCTATCAGCAAGTATGAGCTACAGAAACCGTGCACAAAGACTTCTACAGAGTG 540  
541 AACCCGAAATCTCTCTCTGAGAGAAACAGTCTAGTGTCAACCTCTCAACCTTGGAA 600  
541 AACCCGAAATCTCTCTCTGAGAGAAACAGTCTAGTGTCAACCTCTCAACCTTGGAA 600  
601 CTGAGAGAGCTGAGAGCAAGAGGAGATACCAACCTCAACAGCTCTCTACAGATG 660  
601 CTGAGAGAGCTGAGAGCAAGAGGAGATACCAACCTCAACAGCTCTCTACAGATG 660  
661 AATTGGATCTGATTTCTTCTGAGAGAAACAGTCTAGTGTCAACCTCTCAACCTTGGAA 720  
661 AATTGGATCTGATTTCTTCTGAGAGAAACAGTCTAGTGTCAACCTCTCAACCTTGGAA 720  
721 ATCAAGAAATTTTCAAAATCAACCCCTCAAGAGACCGAGGATGAATCACTGTTGATTCG 780  
721 ATCAAGAAATTTTCAAAATCAACCCCTCAAGAGACCGAGGATGAATCACTGTTGATTCG 780  
781 CAAAAAGAGCTGCTGATTTTCTGAGAGAGATGAATCACTGTTGATTCG 840  
781 CAAAAAGAGCTGCTGATTTTCTGAGAGAGATGAATCACTGTTGATTCG 840  
841 CCAAGTAATATGATTTTGAACCACTGAGAGCGTGCAGTGAAGGATCCAGAAAAAGT 900  
841 CCAAGTAATATGATTTTGAACCACTGAGAGCGTGCAGTGAAGGATCCAGAAAAAGT 900  
901 ATCAAGAGATTTTCAAAATCAACCCCTCAAGAGACCGAGGATGAATCACTGTTGATTCG 960  
901 ATCAAGAGATTTTCAAAATCAACCCCTCAAGAGACCGAGGATGAATCACTGTTGATTCG 960  
961 GCTATTTACAGCATGAGAGCAAGAGCTTATGATCACTGAAGAGAGATGAATGATAGAA 1020  
961 GCTATTTACAGCATGAGAGCAAGAGCTTATGATCACTGAAGAGAGATGAATGATAGAA 1020  
1021 AGGCTGAATTTCTGATTAATAAGCAAGAGCTGCTTACAGAGAGGATCAATCAAGAT 1080  
1021 AGGCTGAATTTCTGATTAATAAGCAAGAGCTGCTTACAGAGAGGATCAATCAAGAT 1080  
1081 GGGCTGGAAGTAAAGCAATGATATGAGCGGAGCTCCAGACAGAAAAAGGATG 1140  
1081 GGGCTGGAAGTAAAGCAATGATATGAGCGGAGCTCCAGACAGAAAAAGGATG 1140  
1141 ATCTGAAGTGTGATCCCTGTGTGAGAGAAAGATGAATGAAGAACTGCAATGCT 1200  
1141 ATCTGAAGTGTGATCCCTGTGTGAGAGAAAGATGAATGAAGAACTGCAATGCT 1200

1201 CAGAGAAATCTAGAGATACAGAGATGCTTCTGAGATTAACCTAATAGCAGATTCAGA 1260  
1201 CAGAGAAATCTAGAGATACAGAGATGCTTCTGAGATTAACCTAATAGCAGATTCAGA 1260  
1261 AAGTTAATGAGTGTGTTCCAGAGAGTGTGAACCTTTAGTTGCTGATACACATGATG 1320  
1261 AAGTTAATGAGTGTGTTCCAGAGAGTGTGAACCTTTAGTTGCTGATACACATGATG 1320  
1321 GGGAGTGTGATTAACCTCAAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1380  
1321 GGGAGTGTGATTAACCTCAAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1380  
1381 AATATTTCTGTTCTTCAAGAGAAATAGACTTACAGGAGATGATGATGATGATGATG 1440  
1381 AATATTTCTGTTCTTCAAGAGAAATAGACTTACAGGAGATGATGATGATGATGATG 1440  
1441 TATGTAAGAGTGAAGAGTCACTTCAATCAGTGAAGATTAATTTGAAGCAAAATAT 1500  
1441 TATGTAAGAGTGAAGAGTCACTTCAATCAGTGAAGATTAATTTGAAGCAAAATAT 1500  
1501 TTGGGAAAGCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
1501 TTGGGAAAGCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
1561 TAAATTTAGAGCATTTTGTACTGAGCAGATTAATCAAGAGGCTCCCTCACAATA 1620  
1561 TAAATTTAGAGCATTTTGTACTGAGCAGATTAATCAAGAGGCTCCCTCACAATA 1620  
1621 AATTAAAGCGTAAAGAGAGAGCTTACATGAGGCTTCACTGAGAGATTTATCAAGAA 1680  
1621 AATTAAAGCGTAAAGAGAGAGCTTACATGAGGCTTCACTGAGAGATTTATCAAGAA 1680  
1681 CAGATTTGGAGTCAAAAGAGAGCTCTGTAATGATTAATCAAGAGAGTCAACCAAGG 1740  
1681 CAGATTTGGAGTCAAAAGAGAGCTCTGTAATGATTAATCAAGAGAGTCAACCAAGG 1740  
1741 AGAATGTCAGATGATTAATTAATTAATGATGATGATGATGATGATGATGATGATG 1800  
1741 AGAATGTCAGATGATTAATTAATTAATGATGATGATGATGATGATGATGATGATG 1800  
1801 CTATTTCAAGATGAGAAAAATCTTAACCCATGAATCACTGTAAGAAATCTGTTTCA 1860  
1801 CTATTTCAAGATGAGAAAAATCTTAACCCATGAATCACTGTAAGAAATCTGTTTCA 1860  
1861 AAAGGAAAGCTGAGAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
1861 AAAGGAAAGCTGAGAGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
1921 ACAATTTCAAAAGCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 1980  
1921 ACAATTTCAAAAGCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 1980  
1981 ATGCGCTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
1981 ATGCGCTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
2041 TTGATAGTGTGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
2041 TTGATAGTGTGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
2101 GGCACAG 2160  
2101 GGCACAG 2160  
2161 GTAAACAG 2220  
2161 GTAAACAG 2220  
2221 AGTTAAACAG 2280  
2221 AGTTAAACAG 2280  
2281 TTGTCATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
2281 TTGTCATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340



D	b	2281	TTGTCAATCTTAGCTTCCAAAGAGAGAAAGAAAGAACTAGAAACGTTAAAGTGT	2340
Q	y	2341	CTAATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGGAGAAAGGTTTGGCAACTG	2400
D	b	2341	CTAATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGGAGAAAGGTTTGGCAACTG	2400
Q	y	2401	AAAGATCTGTAGAGAGTACAGATATTTCAATGGTACTGGTACTGATTTATGGCAGT	2460
D	b	2401	AAAGATCTGTAGAGAGTACAGATATTTCAATGGTACTGGTACTGATTTATGGCAGT	2460
Q	y	2461	AAAGATCTGTAGAGAGTACAGATATTTCAATGGTACTGGTACTGATTTATGGCAGT	2520
D	b	2461	AAAGATCTGTAGAGAGTACAGATATTTCAATGGTACTGGTACTGATTTATGGCAGT	2520
Q	y	2521	GTGTAGAGTACAGTGTGAGATTTGAAACCCCAAGGACTAATTCATGGTGTTCGCAAG	2580
D	b	2521	GTGTAGAGTACAGTGTGAGATTTGAAACCCCAAGGACTAATTCATGGTGTTCGCAAG	2580
Q	y	2581	ATTAATGAGAAATGACACAGAGGCTTTAAGTATCCATTTGGACATGAAGTTAACACAGTC	2640
D	b	2581	ATTAATGAGAAATGACACAGAGGCTTTAAGTATCCATTTGGACATGAAGTTAACACAGTC	2640
Q	y	2641	GGGAAACAGAGCTAGAGAAATGAGAAAGAACTTATGCTCAGTATTTTGGCAAAATACAT	2700
D	b	2641	GGGAAACAGAGCTAGAGAAATGAGAAAGAACTTATGCTCAGTATTTTGGCAAAATACAT	2700
Q	y	2701	TCAAGGTTTCAAAAGGCCAGTATTTGCTCTGTCTTCAATCCAGGAAATGCGAGAGAG	2760
D	b	2701	TCAAGGTTTCAAAAGGCCAGTATTTGCTCTGTCTTCAATCCAGGAAATGCGAGAGAG	2760
Q	y	2761	AATGTGCAACATTTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAAGTCACTT	2820
D	b	2761	AATGTGCAACATTTCTGCCCCACTCTGGGTCCTTAAGAAACAAAGTCCAAAAGTCACTT	2820
Q	y	2821	TTGAAATGGAACAAAAGAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCCCTGTAC	2880
D	b	2821	TTGAAATGGAACAAAAGAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCCCTGTAC	2880
Q	y	2881	AGACAGTTAATATTCACAGCTTCTGCTGGTGGTGCAGAAAGATTAAGCCAGTTGATA	2940
D	b	2881	AGACAGTTAATATTCACAGCTTCTGCTGGTGGTGCAGAAAGATTAAGCCAGTTGATA	2940
Q	y	2941	ATGCCAAATGTAGATCAAAAGAGGCTTAGGTTTTGTCTATCATCTCAGTTCCAGAGCA	3000
D	b	2941	ATGCCAAATGTAGATCAAAAGAGGCTTAGGTTTTGTCTATCATCTCAGTTCCAGAGCA	3000
Q	y	3001	ACGAAACTGTGACTTACTTCCAAATTAACATGAGATTTTACAAAACCATATCGTATAC	3060
D	b	3001	ACGAAACTGTGACTTACTTCCAAATTAACATGAGATTTTACAAAACCATATCGTATAC	3060
Q	y	3061	CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATGTAAAGAAATCTGTCTAGAG	3120
D	b	3061	CACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATGTAAAGAAATCTGTCTAGAG	3120
Q	y	3121	AAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAAATGGAATGGAACATTTCCAA	3180
D	b	3121	AAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAAATGGAATGGAACATTTCCAA	3180
Q	y	3181	GTACAGTGTGACCAATTTAGCCGTAATACATTTAGAGAAATGTTTTTAAAGAACCCAGCT	3240
D	b	3181	GTACAGTGTGACCAATTTAGCCGTAATACATTTAGAGAAATGTTTTTAAAGAACCCAGCT	3240
Q	y	3241	CAAGCAATTTAATGAGAGTGTCCAGTACTAATGAATGGGCTCCATATTTAATGAAA	3300
D	b	3241	CAAGCAATTTAATGAGAGTGTCCAGTACTAATGAATGGGCTCCATATTTAATGAAA	3300
Q	y	3301	TAGGTTCCAGTGTATGAAAACATTTCAAGGAGACTAGTGAACAGAGGCCCAAAATTTGA	3360
D	b	3301	TAGGTTCCAGTGTATGAAAACATTTCAAGGAGACTAGTGAACAGAGGCCCAAAATTTGA	3360
Q	y	3361	ATGCTATGCTTAGATTAGGGGTTTTGGCACTGAGGCTTATTAACAAAGTCTTCTGGA	3420
D	b	3361	ATGCTATGCTTAGATTAGGGGTTTTGGCACTGAGGCTTATTAACAAAGTCTTCTGGA	3420



```

QY 4501 GTGAATACCTATZAGCCAGATTCAGAGGCCCTTTGCTGACAACTTTGAGGTGCTG 4560
    |||||||
Db 4501 GTGAATACCTATZAGCCAGATTCAGAGGCCCTTTGCTGACAACTTTGAGGTGCTG 4560
QY 4561 CAGATAGTCTTACCAGTAAATAAAGAACCCAGAGTGAAGGTCATCCCTCTTAAT 4620
    |||||||
Db 4561 CAGATAGTCTTACCAGTAAATAAAGAACCCAGAGTGAAGGTCATCCCTCTTAAT 4620
QY 4621 GCCCATCATTAAGTATGAGTGTGACATGACAGTTGCTGAGTCTTACAGATGA 4680
    |||||||
Db 4621 GCCCATCATTAAGTATGAGTGTGACATGACAGTTGCTGAGTCTTACAGATGA 4680
QY 4681 ACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGCAACAGCTGGA 4740
    |||||||
Db 4681 ACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGCAACAGCTGGA 4740
QY 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTCCAGAGCAAGATCTAGAG 4800
    |||||||
Db 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTCCAGAGCAAGATCTAGAG 4800
QY 4801 CCCCTTACCTGGAATCTGGAATCAGCCCTCTCTGATGACCCGAAATCTGATCT 4860
    |||||||
Db 4801 CCCCTTACCTGGAATCTGGAATCAGCCCTCTCTGATGACCCGAAATCTGATCT 4860
QY 4861 AAGACAGAGCCCGAGAGTCACTGCTGTGCAACATACCATCTTCAACCTCTGAT 4920
    |||||||
Db 4861 AAGACAGAGCCCGAGAGTCACTGCTGTGCAACATACCATCTTCAACCTCTGAT 4920
QY 4921 AAGTCCCAATTAAGAGTGTGCAAGATCTGCCAGAGTCCAGCTGCTCATACT 4980
    |||||||
Db 4921 AAGTCCCAATTAAGAGTGTGCAAGATCTGCCAGAGTCCAGCTGCTCATACT 4980
QY 4981 ATACTGCTGGTATTAATGCAATGGAAGTGTGACAGAGCAACCCGAATGTGAC 5040
    |||||||
Db 4981 ATACTGCTGGTATTAATGCAATGGAAGTGTGACAGAGCAACCCGAATGTGAC 5040
QY 5041 CTTCAACAGAAAGGTCACAAAAGAAATGTCATGAGTGTGCTGAGCTACCCAGA 5100
    |||||||
Db 5041 CTTCAACAGAAAGGTCACAAAAGAAATGTCATGAGTGTGCTGAGCTACCCAGA 5100
QY 5101 AATTATGCTGCTGATACAGTTGGCAGAAAACACACATCACTTAACTAATTA 5160
    |||||||
Db 5101 AATTATGCTGCTGATACAGTTGGCAGAAAACACACATCACTTAACTAATTA 5160
QY 5161 CTGAAGAGTACTCATCTGATGTAAGAAACAGATGCTGAGTGTGTAACGAGAC 5220
    |||||||
Db 5161 CTGAAGAGTACTCATCTGATGTAAGAAACAGATGCTGAGTGTGTAACGAGAC 5220
QY 5221 TGAATATTTTCTAGAAATGCGGAGGAAATGGTAGTATTTCTGGGTGACC 5280
    |||||||
Db 5221 TGAATATTTTCTAGAAATGCGGAGGAAATGGTAGTATTTCTGGGTGACC 5280
QY 5281 AGTCTATTAAGAAAGAAATGCTGATGAGCATGATTTTGAAGTCAAGAGATGTG 5340
    |||||||
Db 5281 AGTCTATTAAGAAAGAAATGCTGATGAGCATGATTTTGAAGTCAAGAGATGTG 5340
QY 5341 TCAATGGAAGAAACCCCAAGGTCGAAGGAGCAAGATGCCAGAGAAAGATCT 5400
    |||||||
Db 5341 TCAATGGAAGAAACCCCAAGGTCGAAGGAGCAAGATGCCAGAGAAAGATCT 5400
QY 5401 TCAGAGGGGTAGAAATGTTGCTATGAGGCTTACACACATGCCAGAGATCAATG 5460
    |||||||
Db 5401 TCAGAGGGGTAGAAATGTTGCTATGAGGCTTACACACATGCCAGAGATCAATG 5460
QY 5461 AATGATGATACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5520
    |||||||
Db 5461 AATGATGATACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5520
QY 5521 GCACAGGTGTCACCCCAATGTTGTTGTGAGCCAGATGCTGAGAGAGCAATGGCT 5580
    |||||||
Db 5521 GCACAGGTGTCACCCCAATGTTGTTGTGAGCCAGATGCTGAGAGAGCAATGGCT 5580

```

```

QY 5581 TCCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCGAGAGTGGGTGGACA 5640
    |||||||
Db 5581 TCCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCGAGAGTGGGTGGACA 5640
QY 5641 GTTAGACACTTACCAGTGTGCGAGAGCTGGACACTTACTATATACCCAGATGCCCCACA 5700
    |||||||
Db 5641 GTTAGACACTTACCAGTGTGCGAGAGCTGGACACTTACTATATACCCAGATGCCCCACA 5700
QY 5701 GCCACTACTGA 5711
    |||||||
Db 5701 GCCACTACTGA 5711

RESULT 8
AAV46459
ID AAV46459 standard; cDNA: 5711 BP.
XX
AC AAV46459;
XX
DT 18-NOV-1998 (first entry)
XX
DE Human BRCA1 omi2 polymorphism #2 cDNA.
XX
KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
KW chromosome 17q; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT /*tag= a
FT /product= "BRCA1 omi2 protein"
FT variation 2430
FT /*tag= b
FT /note= "This polymorphic variation can be a T or C
FT nucleotide"
XX
PN US5750400-A.
XX
PD 12-MAY-1998.
XX
PF 12-FEB-1997; 97US-0798691.
XX
PR 12-FEB-1996; 96US-0598591.
PR 12-FEB-1997; 97US-0798691.
XX
PA (ONCO-) ONCOMED INC.
XX
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
PI Scheller DB, Zeng B;
XX
DR WPI: 1998-296774/26.
XX
PT BRCA1 omi gene coding sequences - useful for distinguishing between
PT polymorphisms and mutation(s) in the screening for disposition to
PT breast or ovarian cancer
XX
PS Claim 2e; Page -: 54pp; English.
XX
CC This sequence encodes a human BRCA1 (breast and ovarian cancer
CC predisposing gene) omi2 gene in which a polymorphic variation occurs at
CC nucleotide 2430. This sequence and other polymorphic variations of this
CC sequence are useful for the identification of an individual who may or
CC may not have an increased susceptibility to breast or ovarian cancer.
CC The sequences used identify gene changes which are due to polymorphisms,
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC suppressor) which is involved in genetic inheritance of cancers,
CC especially breast and ovarian cancer. It is found at human chromosome
CC 17q which is known to be linked to cancer susceptibility, especially
CC breast cancer. Cells containing a mutation in this gene lose the
CC wild-type function of BRCA1 and are more susceptible to cancers.
CC NOTE: This sequence does not appear in the specification but has been

```

CC created from the wild type BRCA1 omi2 gene represented in AAV46449.  
XX  
Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;  
Query Match 100.0%; Score 5709; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCTCGGTGAGACTCTCTGGACCCCGCAGCAGGCTGGGGTTCTCAGATAACTGGGCC 60  
DB 1 AGCTCGGTGAGACTCTCTGGACCCCGCAGCAGGCTGGGGTTCTCAGATAACTGGGCC 60

QY 61 CCGCGCTCAGAGAGGCTTCAACCTCTCTGGGTAAGTTCAATGGAACAGAGAAA 120  
DB 61 CCGCGCTCAGAGAGGCTTCAACCTCTCTGGGTAAGTTCAATGGAACAGAGAAA 120

QY 121 TGGATTATCTGCTCTCGGGTTGAAGAAGTAAATGATTAATGCTATGAGAAA 180  
DB 121 TGGATTATCTGCTCTCGGGTTGAAGAAGTAAATGATTAATGCTATGAGAAA 180

QY 181 TCTTAGAGTGTCCATCTCTGTGAGTTGATCAAGAACTGTCTCCACAAGTGTGACC 240  
DB 181 TCTTAGAGTGTCCATCTCTGTGAGTTGATCAAGAACTGTCTCCACAAGTGTGACC 240

QY 241 ACATATTTTGGCAAAATTTTGCATGCTGAACCTTCAACACAGAGAAAGGCTTCAACAGT 300  
DB 241 ACATATTTTGGCAAAATTTTGCATGCTGAACCTTCAACACAGAGAAAGGCTTCAACAGT 300

QY 301 GTCCCTTAGTAAATGATGATTAACCAAAAGAGCCCTCAAGAAAGTACAGATTTAGTC 360  
DB 301 GTCCCTTAGTAAATGATGATTAACCAAAAGAGCCCTCAAGAAAGTACAGATTTAGTC 360

QY 361 AACTGTTGAAGAGCTATTGAAATCAATTTGTGCTTTACGCTTGACAGAGTTGGAGT 420  
DB 361 AACTGTTGAAGAGCTATTGAAATCAATTTGTGCTTTACGCTTGACAGAGTTGGAGT 420

QY 421 ATGCAAAACAGCTAATTTTGGCAAAAAGAAATTAATCTCTGAAACATCTAAAGATG 480  
DB 421 ATGCAAAACAGCTAATTTTGGCAAAAAGAAATTAATCTCTGAAACATCTAAAGATG 480

QY 481 AAGTTCTATCATCAAGATGATGGGCTACAGAAACGTCGCAAAAGACTCTACAGAGT 540  
DB 481 AAGTTCTATCATCAAGATGATGGGCTACAGAAACGTCGCAAAAGACTCTACAGAGT 540

QY 541 AACCCGAAATCCCTCTCTGAGAAACAGCTCTCACTGTCACAACTCTTAACCTTGAA 600  
DB 541 AACCCGAAATCCCTCTCTGAGAAACAGCTCTCACTGTCACAACTCTTAACCTTGAA 600

QY 601 CTGTGGAAGCTCTGAGGACAAACAGGAGTACAACTCAAAAGACCTGTCTACATTTG 660  
DB 601 CTGTGGAAGCTCTGAGGACAAACAGGAGTACAACTCAAAAGACCTGTCTACATTTG 660

QY 661 AATGAGATCTGATCTCTGGAAGTACCTTAATAGGCACTTATGAGTGGAG 720  
DB 661 AATGAGATCTGATCTCTGGAAGTACCTTAATAGGCACTTATGAGTGGAG 720

QY 721 ATCAGAAATTTGTTCAAAATCACCCCTCAAGAAACAGGAGTGAATCAATTTGGATCTG 780  
DB 721 ATCAGAAATTTGTTCAAAATCACCCCTCAAGAAACAGGAGTGAATCAATTTGGATCTG 780

QY 781 CAAAAAAGGCTGCTTGAATTTTCTGAGACGATGAACAAATACGACATCATCAAC 840  
DB 781 CAAAAAAGGCTGCTTGAATTTTCTGAGACGATGAACAAATACGACATCATCAAC 840

QY 841 CGAGTAATATGATTTGAACACACGATGAGAGGTCAGCTGAGAGCATCCAGAAAGT 900  
DB 841 CGAGTAATATGATTTGAACACACGATGAGAGGTCAGCTGAGAGCATCCAGAAAGT 900

QY 901 ATCAGAGTATGTTCTGTTTCAAACTTGATGAGGCTATGACACAAATACATCAATGCCA 960  
DB 901 ATCAGAGTATGTTCTGTTTCAAACTTGATGAGGCTATGACACAAATACATCAATGCCA 960

QY 961 GCTCATATACAGATGAGAAACAGCAGTTTATTACTCACTAAAGACAGATGATGTAGAAA 1020

DB 961 GCTCATATACAGATGAGAAACAGCAGTTTATTACTCACTAAAGACAGAAATGATAGAAA 1020

QY 1021 AGCTGAATCTGTAATTAANAAGCAACAGGCTGGCTTAGACAGAGCCAAATACAGAT 1080  
DB 1021 AGCTGAATCTGTAATTAANAAGCAACAGGCTGGCTTAGACAGAGCCAAATACAGAT 1080

QY 1081 GGGCTGGAAGTAAAGCAACATGTAATGATAGGCGGACCTCCACACAGAAAAAAGGTAG 1140  
DB 1081 GGGCTGGAAGTAAAGCAACATGTAATGATAGGCGGACCTCCACACAGAAAAAAGGTAG 1140

QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAAGCAGAAATGCCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAAGCAGAAATGCCATGCT 1200

QY 1201 CAGGAATCCCTAGAGATGATGAGATGCTCTGGATGATTAACACTAAATATACAGATTCACA 1260  
DB 1201 CAGGAATCCCTAGAGATGATGAGATGCTCTGGATGATTAACACTAAATATACAGATTCACA 1260

QY 1261 AAGTTAATGATGATGTTTCCAGAAGTATGATGATGATGATGATGATGATGATGATGATG 1320  
DB 1261 AAGTTAATGATGATGTTTCCAGAAGTATGATGATGATGATGATGATGATGATGATGATG 1320

QY 1321 GGGAGTCTGAATCAATGCCAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 GGGAGTCTGAATCAATGCCAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1380

QY 1381 AATATCTGTTCTTCCAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1381 AATATCTGTTCTTCCAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1440

QY 1441 TATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
DB 1441 TATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500

QY 1501 TTGGGAAACCTATCGGAGAGGCAAGCTCCCACTTAAGCCATGTAACGTAAGAAATC 1560  
DB 1501 TTGGGAAACCTATCGGAGAGGCAAGCTCCCACTTAAGCCATGTAACGTAAGAAATC 1560

QY 1561 TAATTAATGAGAGATTTGTTACTGAGCCAGATTAATGATGATGATGATGATGATGATGATGAT 1620  
DB 1561 TAATTAATGAGAGATTTGTTACTGAGCCAGATTAATGATGATGATGATGATGATGATGATGAT 1620

QY 1621 AATTAAGCGTAAAGAGACCTACATCAAGGCTTCATCCTGAGATTTTATCAAGAAAG 1680  
DB 1621 AATTAAGCGTAAAGAGACCTACATCAAGGCTTCATCCTGAGATTTTATCAAGAAAG 1680

QY 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAGACTTAACCAACGAGAC 1740  
DB 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAGACTTAACCAACGAGAC 1740

QY 1741 AGAATGCTCAAGTATGATTAATTAATAGGTCATGAGATTAATTAATTAATTAATTAATTAAT 1800  
DB 1741 AGAATGCTCAAGTATGATTAATTAATAGGTCATGAGATTAATTAATTAATTAATTAATTAAT 1800

QY 1801 CTATTCAGATGAGAAATTCCTAACCCATAGAAATCACTCCAGAAAGATCTGCTTTCA 1860  
DB 1801 CTATTCAGATGAGAAATTCCTAACCCATAGAAATCACTCCAGAAAGATCTGCTTTCA 1860

QY 1861 AAACGAAAGCTGAACTATTAAGCAGAGTAAAGCAATTTGGAACCTGCAATTAATATCC 1920  
DB 1861 AAACGAAAGCTGAACTATTAAGCAGAGTAAAGCAATTTGGAACCTGCAATTAATATCC 1920

QY 1921 ACAATTTCAAAAGCACTTAAGAAATAGGCTGAGGAGGAGTCTTCTACAGGCAATATTC 1980  
DB 1921 ACAATTTCAAAAGCACTTAAGAAATAGGCTGAGGAGGAGTCTTCTACAGGCAATATTC 1980

QY 1981 ATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040  
DB 1981 ATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040

QY 2041 TTGATAGTGTCTTACAGTGAAGAGATTAAGAAAAAAGTACCAACCAATGCCAGTCA 2100  
DB 2041 TTGATAGTGTCTTACAGTGAAGAGATTAAGAAAAAAGTACCAACCAATGCCAGTCA 2100

Db 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAAAGTACACCAAAATGCCACTCA 2100  
Qy 2101 GGCACAGCAGAAACCTACAACTCATGGAAGTAAAGAACTGCATCTGAGCACAAGAGA 2160  
Db 2101 GGCACAGCAGAAACCTACAACTCATGGAAGTAAAGAACTGCATCTGAGCACAAGAGA 2160  
Qy 2161 GTAACAAGCCAAATGAACAGACAGATTAAGAACATGACAGTACTTTCCACAGAGCTGA 2220  
Db 2161 GTAACAAGCCAAATGAACAGACAGATTAAGAACATGACAGAGTACTTTCCACAGAGCTGA 2220  
Qy 2221 AGTTAAACAAATGCACCTGGTCTTTCTTACTAAGTCTTAACATACGAGTAAAGAT 2280  
Db 2221 AGTTAAACAAATGCACCTGGTCTTTCTTACTAAGTCTTAACATACGAGTAAAGAT 2280  
Qy 2281 TTGCAATCTAGCCTTCCAGAGAGAGAAAAAGAAAGAACTGAAGAACTTAAGTGT 2340  
Db 2281 TTGCAATCTAGCCTTCCAGAGAGAGAAAAAGAAAGAACTGAAGAACTTAAGTGT 2340  
Qy 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGCACACTG 2400  
Db 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGCACACTG 2400  
Qy 2401 AAAATCTGTAGAGATGACAGTATTTTCATTTGCTGCTGCTGCTGCTGCTGCTGCTG 2460  
Db 2401 AAAATCTGTAGAGATGACAGTATTTTCATTTGCTGCTGCTGCTGCTGCTGCTGCTG 2460  
Qy 2461 AAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520  
Db 2461 AAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520  
Qy 2521 GTGTGAGTCACTGTGACAGCATTTTGAACCCCAAGGAGCTAATTCATGTTTGCACAAAG 2580  
Db 2521 GTGTGAGTCACTGTGACAGCATTTTGAACCCCAAGGAGCTAATTCATGTTTGCACAAAG 2580  
Qy 2581 ATATAGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGTC 2640  
Db 2581 ATATAGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGTC 2640  
Qy 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700  
Db 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700  
Qy 2701 TCAAGTTTCAAGGCGCCAGTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760  
Db 2701 TCAAGTTTCAAGGCGCCAGTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760  
Qy 2761 AATGTGCAACATTTCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
Db 2761 AATGTGCAACATTTCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
Qy 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGATGAGTCAATATCAAGCCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAAAGAAATCAAGAAAGATGAGTCAATATCAAGCCTGTAC 2880  
Qy 2881 AGACAGTTAATATCATCTGAGGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940  
Db 2881 AGACAGTTAATATCATCTGAGGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940  
Qy 2941 ATGCGCAATGTAGTATCAAGAGAGGCTGTAGGTTTGTCTATCTGCTGCTGCTGCTGCTG 3000  
Db 2941 ATGCGCAATGTAGTATCAAGAGAGGCTGTAGGTTTGTCTATCTGCTGCTGCTGCTGCTG 3000  
Qy 3001 ACGAAATCTGACATCTACTGCTCAATTAACATGAGTCAATTAACATGAGTCAATGATAC 3060  
Db 3001 ACGAAATCTGACATCTACTGCTCAATTAACATGAGTCAATTAACATGAGTCAATGATAC 3060  
Qy 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAATGTAAGAAAAATCTGCTAGAG 3120  
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAATGTAAGAAAAATCTGCTAGAG 3120  
Qy 3121 AAAAATTTGAGAGATTCATGCTGCTGAGAGAGAAATGAGAGAAATGAGAGAAATGAGAG 3180  
Db 3121 AAAAATTTGAGAGATTCATGCTGCTGAGAGAGAAATGAGAGAAATGAGAGAAATGAGAG 3180

Qy 3181 GTACAGTGAACATTAAGCCGTAAATTAACATTAGAGAAATGTTTTAAAGACCCAGCT 3240  
Db 3181 GTACAGTGAACATTAAGCCGTAAATTAACATTAGAGAGAAATGTTTTAAAGACCCAGCT 3240  
Qy 3241 CAAGCAATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCGATTAATGAAG 3300  
Db 3241 CAAGCAATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCGATTAATGAAG 3300  
Qy 3301 TAGGTTCCAGTGAAGAAACATTCAGACAGAACTAGTGAAGAAACAGAGGCGCAAAATGA 3360  
Db 3301 TAGGTTCCAGTGAAGAAACATTCAGACAGAACTAGTGAAGAAACAGAGGCGCAAAATGA 3360  
Qy 3361 ATGCTATGCTAGATTAAGTGGGTTTTCAGCTGAGGCTGATTAACAAAGCTTCTCTGAA 3420  
Db 3361 ATGCTATGCTAGATTAAGTGGGTTTTCAGCTGAGGCTGATTAACAAAGCTTCTCTGAA 3420  
Qy 3421 GTAATGTAAACATCTGTAATAAAGCAAGAAATGAAGAGTGAAGTGAAGTGAAGTGA 3480  
Db 3421 GTAATGTAAACATCTGTAATAAAGCAAGAAATGAAGAGTGAAGTGAAGTGAAGTGA 3480  
Qy 3481 ATACAGATTTCTCTCATATCTGATTTTCAGATTAACATTGAACACCTATGGAGTATG 3540  
Db 3481 ATACAGATTTCTCTCATATCTGATTTTCAGATTAACATTGAACACCTATGGAGTATG 3540  
Qy 3541 ATGATCTCAAGTGTGTTCTGAGACACCTGATGACCTGTTGATGATGTTGAATTAAG 3600  
Db 3541 ATGATCTCAAGTGTGTTCTGAGACACCTGATGACCTGTTGATGATGTTGAATTAAG 3600  
Qy 3601 AAGATACTAGTTTGTCTGAAATGACATTAAGAAAGTTCGCTGTTTTAGCAAAACCG 3660  
Db 3601 AAGATACTAGTTTGTCTGAAATGACATTAAGAAAGTTCGCTGTTTTAGCAAAACCG 3660  
Qy 3661 TCCAGAAAGGAGACTTAAGCAGAGAGTCTTACCCCTTACCCATACACATTGGCTCAGG 3720  
Db 3661 TCCAGAAAGGAGACTTAAGCAGAGAGTCTTACCCCTTACCCATACACATTGGCTCAGG 3720  
Qy 3721 GTTACCGAAGAGGAGGAGCAAGAAATTAAGTCTGAGAGAACTTATCTGATGAGATG 3780  
Db 3721 GTTACCGAAGAGGAGGAGCAAGAAATTAAGTCTGAGAGAACTTATCTGATGAGATG 3780  
Qy 3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTTAAGTAAACAAATATACCTTCTCAGT 3840  
Db 3781 AAGAGCTTCCCTGCTTCCACACCTGTTATTTGTTAAGTAAACAAATATACCTTCTCAGT 3840  
Qy 3841 CTACTAGGCAATGACACCGTCTGCTACCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3900  
Db 3841 CTACTAGGCAATGACACCGTCTGCTACCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3900  
Qy 3901 TATCATTGAAGAAATAGCTTAAATGACCTGCACTAACACAGTAAATTTGGCAAAAGCATCTC 3960  
Db 3901 TATCATTGAAGAAATAGCTTAAATGACCTGCACTAACACAGTAAATTTGGCAAAAGCATCTC 3960  
Qy 3961 AGGAACATCACCTTAAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCCACAGTGA 4020  
Db 3961 AGGAACATCACCTTAAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCCACAGTGA 4020  
Qy 4021 GTGAAATTTGAAGACTTGAAGTCAAAATACAAACACCGAGATCTTCTTGAATTTGCTCT 4080  
Db 4021 GTGAAATTTGAAGACTTGAAGTCAAAATACAAACACCGAGATCTTCTTGAATTTGCTCT 4080  
Qy 4081 CCAACAAATGAGGCACTGCTGTAAGAGCCAGGAGTGTGCTGAGTGAACAAAGAAATTTGG 4140  
Db 4081 CCAACAAATGAGGCACTGCTGTAAGAGCCAGGAGTGTGCTGAGTGAACAAAGAAATTTGG 4140  
Qy 4141 TTTTCAGATGATGAAGAAAGAGAGAGGCTTGAAGAAATTAATCAAGAAAGACCAAGCA 4200  
Db 4141 TTTTCAGATGATGAAGAAAGAGAGAGGCTTGAAGAAATTAATCAAGAAAGACCAAGCA 4200  
Qy 4201 TGGATTTCAAACTTATAGTGAAGAGAGATCTGGGTGTGAGAGTGAAGAAAGAGCTCTCTAG 4260  
Db 4201 TGGATTTCAAACTTATAGTGAAGAGAGATCTGGGTGTGAGAGTGAAGAAAGAGCTCTCTAG 4260

OY	4261	ACTGCTCAGGGCTATCCTCTCAGAGTGCATTTTAACTCCTCAGCAGAGGGATACCATGC	4320
Db	4261	ACTGCTCAGGGCTATCCTCTCAGAGTGCATTTTAACTCCTCAGCAGAGGGATACCATGC	4320
OY	4321	AACATACCTGATTAAGCTCCAGAGAAATGGCTGACTAGAAAGCTGTGTAGACAGC	4380
Db	4321	AACATACCTGATTAAGCTCCAGAGAAATGGCTGACTAGAAAGCTGTGTAGACAGC	4380
OY	4381	ATGGAGCCGACCTTCTAACAGTACCTCCATCATTAAGTACTCTTGCCCTTGAGG	4440
Db	4381	ATGGAGCCGACCTTCTAACAGTACCTCCATCATTAAGTACTCTTGCCCTTGAGG	4440
OY	4441	ACCTCGAATTCAGAACCAAGACATCAGAAAAAGCAGTATTAACTTCAGAAAAATA	4500
Db	4441	ACCTCGAATTCAGAACCAAGACATCAGAAAAAGCAGTATTAACTTCAGAAAAATA	4500
OY	4501	GTAATATACCTTAATAGCAAAATCCAGAAAGCCTTTCTGCTGACAACTTTGAGGTCTG	4560
Db	4501	GTAATATACCTTAATAGCAAAATCCAGAAAGCCTTTCTGCTGACAACTTTGAGGTCTG	4560
OY	4561	CAGATAGTTCTACAGTAAATAAAGAACAGAGTGAAGAGTCAATCCCTCTAAAT	4620
Db	4561	CAGATAGTTCTACAGTAAATAAAGAACAGAGTGAAGAGTCAATCCCTCTAAAT	4620
OY	4621	GCCCATCATTTAGATGATAGGTGACATGCACAGTTGCTCTGGAGTCTTCAGANTAGAA	4680
Db	4621	GCCCATCATTTAGATGATAGGTGACATGCACAGTTGCTCTGGAGTCTTCAGANTAGAA	4680
OY	4681	ACTACCATCTCAAGAGGAGCTCATTAAGTTGTGTGAGGAGGACCAACAGCTGGAAG	4740
Db	4681	ACTACCATCTCAAGAGGAGCTCATTAAGTTGTGTGAGGAGGACCAACAGCTGGAAG	4740
OY	4741	AGTCTGGCCACACGATTTTGACGAAACATCTTACTTGCCAGGCAAGATCTAGAGGAA	4800
Db	4741	AGTCTGGCCACACGATTTTGACGAAACATCTTACTTGCCAGGCAAGATCTAGAGGAA	4800
OY	4801	CCCCCTTACCTGATCTGGAATCAGCCTCTCTGTGATGACCTGAAATCTGATCTCTG	4860
Db	4801	CCCCCTTACCTGATCTGGAATCAGCCTCTCTGTGATGACCTGAAATCTGATCTCTG	4860
OY	4861	AAGACAGAGCCCCAGAGTCTGTTGGCAACATACCATCTTCAACCTCTGCATTTGA	4920
Db	4861	AAGACAGAGCCCCAGAGTCTGTTGGCAACATACCATCTTCAACCTCTGCATTTGA	4920
OY	4921	AAATTCCTCCAAATTTGAAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTG	4980
Db	4921	AAATTCCTCCAAATTTGAAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTG	4980
OY	4981	ATACGTCTGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAACCCGAAATTCAG	5040
Db	4981	ATACGTCTGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAACCCGAAATTCAG	5040
OY	5041	CTTCAACAGAAAGGTCACAAAGAAATGTCATGGTGGTGTGGCTGACCCCGAAG	5100
Db	5041	CTTCAACAGAAAGGTCACAAAGAAATGTCATGGTGGTGTGGCTGACCCCGAAG	5100
OY	5101	AATTTATGCTCGTACAGTTTGCCAGAAACACCAATCACTTAACTTAATCTAATTA	5160
Db	5101	AATTTATGCTCGTACAGTTTGCCAGAAACACCAATCACTTAACTTAATCTAATTA	5160
OY	5161	CTGAAGAGACTACTCATGTTGTATGAAACAGATGCTGAGTTGTGTGAAACGGACAC	5220
Db	5161	CTGAAGAGACTACTCATGTTGTATGAAACAGATGCTGAGTTGTGTGAAACGGACAC	5220
OY	5221	TGAATATTTTCTAGAAATTTGGGAGGAAATGGGAGTATCTATTTTGGGTGACC	5280
Db	5221	TGAATATTTTCTAGAAATTTGGGAGGAAATGGGAGTATCTATTTTGGGTGACC	5280
OY	5281	ACTCTATTTAAAGAAAGAAATGCTGATGAGCATATTTTGAAGTCAGAGGAGATGTG	5340
Db	5281	ACTCTATTTAAAGAAAGAAATGCTGATGAGCATATTTTGAAGTCAGAGGAGATGTG	5340
OY	5341	TCAATGGAAGAAACCAAGAGTCCAAAGCGAGCAAGAGATCCAGGACGAAAGATCT	5400
Db	5341	TCAATGGAAGAAACCAAGAGTCCAAAGCGAGCAAGAGATCCAGGACGAAAGATCT	5400

Db	5341	TCATGGAAGAAACCAAGAGTCCAAAGCGAGCAAGAGATCCAGGACGAAAGATCT	5400
OY	5401	TCAGGGGCTAGAAATCTGTTGCTATGAGGCGCTTTCACCAACATGCCACAGATCAACTG	5460
Db	5401	TCAGGGGCTAGAAATCTGTTGCTATGAGGCGCTTTCACCAACATGCCACAGATCAACTG	5460
OY	5461	AATGATGATGACAGCTGTGTGCTCTGTGTGTGAAAGAGCTTTCATCATTCACCTTG	5520
Db	5461	AATGATGATGACAGCTGTGTGCTCTGTGTGTGAAAGAGCTTTCATCATTCACCTTG	5520
OY	5521	GCAAGGTGTCACCAATTTGTTGTGTCAGCCAGATGCTGAGACAGAGCAATGCT	5580
Db	5521	GCAAGGTGTCACCAATTTGTTGTGTCAGCCAGATGCTGAGACAGAGCAATGCT	5580
OY	5581	TCCATGCAATTTGGGCGAGTGTGAGGACACCTGTGTGACCCGAGAGTGGTGGACA	5640
Db	5581	TCCATGCAATTTGGGCGAGTGTGAGGACACCTGTGTGACCCGAGAGTGGTGGACA	5640
OY	5641	GTTAGACCTCTACAGTGCAGAGCTGACACCTGATGATACCCAGATCCCCACA	5700
Db	5641	GTTAGACCTCTACAGTGCAGAGCTGACACCTGATGATACCCAGATCCCCACA	5700
OY	5701	GCCACTACTGA 5711	
Db	5701	GCCACTACTGA 5711	

RESULT 9  
AAV46460  
ID AAV46460 standard; cDNA; 5711 BP.  
XX  
AC AAV46460;  
XX  
DT 18-NOV-1998 (first entry)  
XX  
DE Human BRCA1 omi2 polymorphism #3 cDNA.  
XX  
KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;  
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
KW chromosome 17q; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT CDS Location/Qualifiers  
FT 120..5711  
FT /\*tag= a  
FT /product= "BRCA1 omi2 protein"  
FT 2731  
FT /\*tag= b  
FT /note= "This polymorphic variation can be a C or T  
nucleotide"  
XX  
PN US5750400-A.  
XX  
PD 12-MAY-1998.  
XX  
PF 12-FEB-1997; 97US-0798691.  
XX  
PR 12-FEB-1996; 96US-0598591.  
PR 12-FEB-1997; 97US-0798691.  
XX  
PA (ONCO-) ONCORMED INC.  
XX  
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Schelter DB, Zeng B;  
XX  
DR WPI; 1998-296774/26.  
XX  
XX BRCA1 omi gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer  
XX

PS Claim 2e; Page -: 54pp; English.  
 CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) om12 gene in which a polymorphic variation occurs at  
 CC nucleotide 2731. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 om12 gene represented in AAV6449.  
 CC  
 XX  
 SQ Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;  
 Query Match 100.0%; Score 5709; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ACCTCGCTGAGACTTCTCTGACCCCGACAGCGCTGGGGTTTTCAGATTAACCTGGCC 60  
 DB 1 AGCTCGCTGAGACTTCTCTGACCCCGACAGCGCTGGGGTTTTCAGATTAACCTGGCC 60  
 QY 61 CCGGCGCTGAGAGGCGCTTACCCCTGCTGGGTAACTGATGAGACAGAAAGAAA 120  
 DB 61 CCGGCGCTGAGAGGCGCTTACCCCTGCTGGGTAACTGATGAGACAGAAAGAAA 120  
 QY 121 TGGATTATTCGCTCTGCGGTTGAAGAAATGCAATATGATGATGAGAGAAA 180  
 DB 121 TGGATTATTCGCTCTGCGGTTGAAGAAATGCAATATGATGATGAGAGAAA 180  
 QY 121 TGGATTATTCGCTCTGCGGTTGAAGAAATGCAATATGATGATGAGAGAAA 180  
 DB 121 TGGATTATTCGCTCTGCGGTTGAAGAAATGCAATATGATGATGAGAGAAA 180  
 QY 181 TCTTAGAGTCCCATCTGCTGAGTGTATCAAGAAAGAACTGCTCCAAAGAGTGACC 240  
 DB 181 TCTTAGAGTCCCATCTGCTGAGTGTATCAAGAAAGAACTGCTCCAAAGAGTGACC 240  
 QY 241 ACATATTTTGAATTTTGGATGCTGAACTGCAACCAAGAAAGGCGCTTCAAGT 300  
 DB 241 ACATATTTTGAATTTTGGATGCTGAACTGCAACCAAGAAAGGCGCTTCAAGT 300  
 QY 241 ACATATTTTGAATTTTGGATGCTGAACTGCAACCAAGAAAGGCGCTTCAAGT 300  
 DB 241 ACATATTTTGAATTTTGGATGCTGAACTGCAACCAAGAAAGGCGCTTCAAGT 300  
 QY 301 GTCTTATATGAGATGATATACCAAAAGAGCCCTACAAAGAAAGTACAGATTATGTC 360  
 DB 301 GTCTTATATGAGATGATATACCAAAAGAGCCCTACAAAGAAAGTACAGATTATGTC 360  
 QY 301 GTCTTATATGAGATGATATACCAAAAGAGCCCTACAAAGAAAGTACAGATTATGTC 360  
 DB 301 GTCTTATATGAGATGATATACCAAAAGAGCCCTACAAAGAAAGTACAGATTATGTC 360  
 QY 361 AACTTTGTGAGAGCTATGAAATCATTTTGTGCTTTTACGTTGACACAGTTTGGAGT 420  
 DB 361 AACTTTGTGAGAGCTATGAAATCATTTTGTGCTTTTACGTTGACACAGTTTGGAGT 420  
 QY 421 ATGCAAAAGAGCTATTTTGGCAAAAAGAAATTAACCTCTGGAACATCTAAAGATG 480  
 DB 421 ATGCAAAAGAGCTATTTTGGCAAAAAGAAATTAACCTCTGGAACATCTAAAGATG 480  
 QY 481 AAGTTTCTATCATCAAAAGATGAGGTACAGAAACCGTGCCAAAGAGCTTCTACAGATG 540  
 DB 481 AAGTTTCTATCATCAAAAGATGAGGTACAGAAACCGTGCCAAAGAGCTTCTACAGATG 540  
 QY 541 AACCCGAAATCTTCTCTGAGAAACCGTCTCAGTGTCCAACTCTAACCTTGGAA 600  
 DB 541 AACCCGAAATCTTCTCTGAGAAACCGTCTCAGTGTCCAACTCTAACCTTGGAA 600  
 QY 601 CTGTGAGAACTGAGAGCAAAAGCAGCGATACACCTCAAAAGAGCTGTCTACATG 660  
 DB 601 CTGTGAGAACTGAGAGCAAAAGCAGCGATACACCTCAAAAGAGCTGTCTACATG 660  
 QY 661 AATTTGAGATCTGATTTCTTGAAGATACGTTAATAGGCACTTATGAGAGTGGAG 720  
 DB 661 AATTTGAGATCTGATTTCTTGAAGATACGTTAATAGGCACTTATGAGAGTGGAG 720  
 QY 721 ATCAAGATTTTACAAATACCCCTCAGAGAAACAGGATGAATCAGTTTGGATTCTG 780  
 DB 721 ATCAAGATTTTACAAATACCCCTCAGAGAAACAGGATGAATCAGTTTGGATTCTG 780

DB 721 ATCAAGATTTTACAAATACCCCTCAGAGAAACAGGATGAATCAGTTTGGATTCTG 780  
 QY 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAAACAAATACTGAAATCATCAAC 840  
 DB 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAAACAAATACTGAAATCATCAAC 840  
 QY 841 CCAATAATATGATTTGAGACACCACTGAGAAAGCGTGCAGCTGAGAGGCAATCAGAAAAAGT 900  
 DB 841 CCAATAATATGATTTGAGACACCACTGAGAAAGCGTGCAGCTGAGAGGCAATCAGAAAAAGT 900  
 QY 901 ATCAGGATGCTGCTTTTCAAACTTGCATGTGAGAGCCATGTGGCACAATATCTATGCCA 960  
 DB 901 ATCAGGATGCTGCTTTTCAAACTTGCATGTGAGAGCCATGTGGCACAATATCTATGCCA 960  
 QY 961 GCTCATTTACAGCATGAGACAGCGATTATTTACTACATAAAGACAGATGAATGTACAAA 1020  
 DB 961 GCTCATTTACAGCATGAGACAGCGATTATTTACTACATAAAGACAGATGAATGTACAAA 1020  
 QY 1021 AGGCTGAATTCCTGTAATTAAGCAACAGCGCTGCTTACAGAGAGCCCAATATACAGAT 1080  
 DB 1021 AGGCTGAATTCCTGTAATTAAGCAACAGCGCTGCTTACAGAGAGCCCAATATACAGAT 1080  
 QY 1081 GGGCTGGAAGTAAAGCAACATGTATATGATAGCGGACTCCAGCACAGAAAAAGGTAG 1140  
 DB 1081 GGGCTGGAAGTAAAGCAACATGTATATGATAGCGGACTCCAGCACAGAAAAAGGTAG 1140  
 QY 1141 ATTCGATGCTGATCCCTGCTGTGAGAAAGAAAGTAAAGCAAACTGCTATGCT 1200  
 DB 1141 ATTCGATGCTGATCCCTGCTGTGAGAAAGAAAGTAAAGCAAACTGCTATGCT 1200  
 QY 1201 CAGAGAAATCTAGAGATCTAGAGATGTTCTTGGATTAACATTAATAGCAGCTTTCAGA 1260  
 DB 1201 CAGAGAAATCTAGAGATCTAGAGATGTTCTTGGATTAACATTAATAGCAGCTTTCAGA 1260  
 QY 1261 AAGTTAATGAGTGGTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
 DB 1261 AAGTTAATGAGTGGTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
 QY 1321 GGGAGTCTGAATCAAAAGTCCAAAGTGCATGATGATGATGATGATGATGATGATGATG 1380  
 DB 1321 GGGAGTCTGAATCAAAAGTCCAAAGTGCATGATGATGATGATGATGATGATGATGATG 1380  
 QY 1381 AATATTCGCTGCTTCTGAGAGAAATAGACTTACTGAGCAATGATCCATGAGGCTTTAA 1440  
 DB 1381 AATATTCGCTGCTTCTGAGAGAAATAGACTTACTGAGCAATGATCCATGAGGCTTTAA 1440  
 QY 1441 TATGTAAAGTGAAGAGTTCCTCAATATCAGTGAAGATTAATTTGAAGACAAATAT 1500  
 DB 1441 TATGTAAAGTGAAGAGTTCCTCAATATCAGTGAAGATTAATTTGAAGACAAATAT 1500  
 QY 1501 TTGGGAAACCTATCGGAAAGAGGAGCTCCCACTTAAGCATGTAATCTGAATATC 1560  
 DB 1501 TTGGGAAACCTATCGGAAAGAGGAGCTCCCACTTAAGCATGTAATCTGAATATC 1560  
 QY 1561 TAAATATAGAGAGATTTTACTGAGAGCAGATTAATTAAGAGAGCGCCCTCACAATA 1620  
 DB 1561 TAAATATAGAGAGATTTTACTGAGAGCAGATTAATTAAGAGAGCGCCCTCACAATA 1620  
 QY 1621 AATTTAAAGCTTAAAGAGAGCTACATCAGAGCTTCAATCTGAGAGATTTATCAAGAA 1680  
 DB 1621 AATTTAAAGCTTAAAGAGAGCTACATCAGAGCTTCAATCTGAGAGATTTATCAAGAA 1680  
 QY 1681 CAGATTTGGAGTTCAGAAAGAGCTGAAATGAAATTCAGGGAACATCAACCAAGGAGC 1740  
 DB 1681 CAGATTTGGAGTTCAGAAAGAGCTGAAATGAAATTCAGGGAACATCAACCAAGGAGC 1740  
 QY 1741 AGAATGGTCAAGTATGATTAATTAATTAATAGTGTGATGAGATTAATAAAGAGTGAT 1800  
 DB 1741 AGAATGGTCAAGTATGATTAATTAATTAATAGTGTGATGAGATTAATAAAGAGTGAT 1800  
 QY 1801 CTATTGAGATGAGAAAAATCTTACCAATCAATCAATCAATCAATCAATCAATCAATCA 1860  
 DB 1801 CTATTGAGATGAGAAAAATCTTACCAATCAATCAATCAATCAATCAATCAATCAATCA 1860



|||||  
 Db 4021 GTGATTTGAGAGCTTACCTGCAAAATACAAACACCCAGGATCTTTCTTGTATGTTCTT 4080  
 OY 4081 CCAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTACAGGAATTCG 4140  
 Db 4081 CCAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTACAGGAATTCG 4140  
 OY 4141 TTTCAGTGTATGAAGAAAGAGGAGGCTTGGAGAAATTAATCAAGAGAGCAAGCA 4200  
 Db 4141 TTTCAGTGTATGAAGAAAGAGGAGGCTTGGAGAAATTAATCAAGAGAGCAAGCA 4200  
 OY 4201 TGGATTCAACTTAGTGAAGACAGCATCTGGTGTGAGTGAAGTGAAGCAAGCGTCTGAG 4260  
 Db 4201 TGGATTCAACTTAGTGAAGACAGCATCTGGTGTGAGTGAAGTGAAGCAAGCGTCTGAG 4260  
 OY 4261 ACTGCTGAGGCTTCTCTCAGAGTACATTTTAACCACTCAGCAGAGGATACCATTC 4320  
 Db 4261 ACTGCTGAGGCTTCTCTCAGAGTACATTTTAACCACTCAGCAGAGGATACCATTC 4320  
 OY 4321 AACATACTGATTAAGCTCCAGCAGGAAATGCTGAAGTGAAGTGTGTAGAACAGC 4380  
 Db 4321 AACATACTGATTAAGCTCCAGCAGGAAATGCTGAAGTGAAGTGTGTAGAACAGC 4380  
 OY 4381 ATGGAGCCAGCCTTCTPAACAGCTACCCCTTCATCATPAAGTACTCTTCCCTTGAAG 4440  
 Db 4381 ATGGAGCCAGCCTTCTPAACAGCTACCCCTTCATCATPAAGTACTCTTCCCTTGAAG 4440  
 OY 4441 ACCTGGCAATCCGAACAAAGCAGCATCAGAAAGAGTATTAACCTTCACAGAAAGTA 4500  
 Db 4441 ACCTGGCAATCCGAACAAAGCAGCATCAGAAAGAGTATTAACCTTCACAGAAAGTA 4500  
 OY 4501 GTGAATACCTTATAGCCAGCAATCCAGAGGCTTCTGCTGACAGTGTGAGAGTCTG 4560  
 Db 4501 GTGAATACCTTATAGCCAGCAATCCAGAGGCTTCTGCTGACAGTGTGAGAGTCTG 4560  
 OY 4561 CAGATAGTCTTACCAAGTAAATTAAGAACCCAGAGTGAAGAGTCAATCCCTTCTAAT 4620  
 Db 4561 CAGATAGTCTTACCAAGTAAATTAAGAACCCAGAGTGAAGAGTCAATCCCTTCTAAT 4620  
 OY 4621 GCCCATATTAGATGAGTGGGTGACATGCAAGTCTGCTGGAGTCTCAGATGAA 4680  
 Db 4621 GCCCATATTAGATGAGTGGGTGACATGCAAGTCTGCTGGAGTCTCAGATGAA 4680  
 OY 4681 ACTACCATCTCAAGAGAGCTCATTAAGTGTGTGATGTGAGAGCAACAGCTGGAAG 4740  
 Db 4681 ACTACCATCTCAAGAGAGCTCATTAAGTGTGTGATGTGAGAGCAACAGCTGGAAG 4740  
 OY 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTGCCAGAGCAAGATCTAGAGGAA 4800  
 Db 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTGCCAGAGCAAGATCTAGAGGAA 4800  
 OY 4801 CCCCTTACCTGGAATCTGGAATCAGGCTCTCTCTGATGAGACCTGATCTGATCTCTG 4860  
 Db 4801 CCCCTTACCTGGAATCTGGAATCAGGCTCTCTCTGATGAGACCTGATCTGATCTCTG 4860  
 OY 4861 AAGACAGAGCCAGAGTCTGCTGTGGAACAATACCATCTTCAACCTCTGCAATGA 4920  
 Db 4861 AAGACAGAGCCAGAGTCTGCTGTGGAACAATACCATCTTCAACCTCTGCAATGA 4920  
 OY 4921 AAGTCCCAATTTGAAGTGTGAGATCTGCCAGAGTCCAGCTGCTGCTCATACTG 4980  
 Db 4921 AAGTCCCAATTTGAAGTGTGAGATCTGCCAGAGTCCAGCTGCTGCTCATACTG 4980  
 OY 4981 ATACTGCTGGGTATTAATGCAATGGAAGTGTGACAGAGGAGAACCCGAATTTGACAG 5040  
 Db 4981 ATACTGCTGGGTATTAATGCAATGGAAGTGTGACAGAGGAGAACCCGAATTTGACAG 5040  
 OY 5041 CTTCAACAGAAAGGCTCAACAAAAGATGTCAATGATGTGTGCTGAGCTTACCCAGAG 5100  
 Db 5041 CTTCAACAGAAAGGCTCAACAAAAGATGTCAATGATGTGTGCTGAGCTTACCCAGAG 5100  
 OY 5101 AATTATGCTGCTGATCAGTTTGCAGAAACACCATCTTAACTTAATCTAATTA 5160  
 Db 5101 AATTATGCTGCTGATCAGTTTGCAGAAACACCATCTTAACTTAATCTAATTA 5160

Db 5101 AATTATGCTGCTGATCAGTTTGCAGAAACACCATCTTAACTTAATCTAATTA 5160  
 OY 5161 CTGAAGAGACTACATGTTTGTATGAAACAGATGCTGATGTTGTGTGACAGCAGC 5220  
 Db 5161 CTGAAGAGACTACATGTTTGTATGAAACAGATGCTGATGTTGTGTGACAGCAGC 5220  
 OY 5221 TGAATATTTTCTAGGAATTCGAGGAGGAAATGGTATGTTAGCTATTTCTGGGTGACC 5280  
 Db 5221 TGAATATTTTCTAGGAATTCGAGGAGGAAATGGTATGTTAGCTATTTCTGGGTGACC 5280  
 OY 5281 AGTCTATTAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGATGTCG 5340  
 Db 5281 AGTCTATTAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGATGTCG 5340  
 OY 5341 TCAATGGAAGAAACCAAGGCTCCAAAGCAGCAAGAGATCCAGAGACGAAGATCT 5400  
 Db 5341 TCAATGGAAGAAACCAAGGCTCCAAAGCAGCAAGAGATCCAGAGACGAAGATCT 5400  
 OY 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCTTCCACCAATGCTCCAGATCACTGG 5460  
 Db 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCTTCCACCAATGCTCCAGATCACTGG 5460  
 OY 5461 AATGATGCTACAGCTGTGTGCTCTCTGCTGTGGAAGAGCTTTCATCATCCCTTG 5520  
 Db 5461 AATGATGCTACAGCTGTGTGCTCTCTGCTGTGGAAGAGCTTTCATCATCCCTTG 5520  
 OY 5521 GCACAGGTGTCACCAATGTTGTTGTCAGCAGCAGATGCTGAGAGAGCAATGGCT 5580  
 Db 5521 GCACAGGTGTCACCAATGTTGTTGTCAGCAGCAGATGCTGAGAGAGCAATGGCT 5580  
 OY 5581 TCCATGCAATTTGGCAGATGTTGAGGACCTGTGTGACCCGAGAGTGGTGTGGAACA 5640  
 Db 5581 TCCATGCAATTTGGCAGATGTTGAGGACCTGTGTGACCCGAGAGTGGTGTGGAACA 5640  
 OY 5641 GTTAGCAGCTTACCAAGTCCAGAGCTGGACACCTTACCTGATCCCAATCCCAACA 5700  
 Db 5641 GTTAGCAGCTTACCAAGTCCAGAGCTGGACACCTTACCTGATCCCAATCCCAACA 5700  
 OY 5701 GCCCAGTCTGCA 5711  
 Db 5701 GCCCAGTCTGCA 5711

RESULT 10  
 AAV46461  
 ID AAV46461 standard; cDNA: 5711 BP.  
 XX AAV46461;  
 AC 18-NOV-1998 (first entry)  
 XX  
 DE Human BRCA1 omi2 polymorphism #4 cDNA.  
 XX  
 KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;  
 KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 OS chromosome 17q; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 120..5711  
 FT CDS /\*tag= a  
 FT /product= "BRCA1 omi2 protein"  
 FT variation 3232  
 FT /\*tag= b  
 FT /note= "This polymorphic variation can be an A or G  
 nucleotide"  
 XX  
 PN US5750400-A.  
 XX  
 PD 12-MAY-1998.  
 XX  
 PF 12-FEB-1997; 97US-0798691.



XX 12-FEB-1996; 96US-0598591.  
 PR 12-FEB-1997; 97US-0798691.  
 XX  
 PA (ONCO-) ONCORMED INC.  
 PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
 PI Schelter DB, Zeng B;  
 XX  
 DR WPI: 1998-296774/26.  
 XX  
 PT BRCA1 omi gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to  
 PT breast or ovarian cancer  
 XX  
 PS  
 XX Claim 2e: Page -: 54pp; English.

CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC predisposing gene) omi2 gene in which a polymorphic variation occurs at  
 CC nucleotide 3232. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 omi2 gene represented in AAV46449.  
 CC  
 SQ Sequence 5711 BP; 1955 A; 1098 C; 1274 G; 1383 T; 1 other;

Query Match 100.0%; Score 5709; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTGCTGAGACTCTCTGACCCCGCAGCAGCTGGGGTTTCTCAGATACTGGCC 60  
 DB 1 AGCTGCTGAGACTCTCTGACCCCGCAGCAGCTGGGGTTTCTCAGATACTGGCC 60  
 OY 61 CCTGCGCTCAGAGGCGCTTACCCCTCTCTCTGGGTTAACTGGAACAGAAAGAA 120  
 DB 61 CCTGCGCTCAGAGGCGCTTACCCCTCTCTCTGGGTTAACTGGAACAGAAAGAA 120  
 OY 121 TGGATTATCTGCTCTCGGCTGAGAGAAATGCAATTAATGATGACAGAAA 180  
 DB 121 TGGATTATCTGCTCTCGGCTGAGAGAAATGCAATTAATGATGACAGAAA 180  
 OY 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
 DB 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
 OY 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
 DB 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
 OY 241 ACATATTTTGGCAAAATTTTGCATCTCTCAACAGAGAAAGGCGCTTACAGT 300  
 DB 241 ACATATTTTGGCAAAATTTTGCATCTCTCAACAGAGAAAGGCGCTTACAGT 300  
 OY 301 GTTCCTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGTACAGATTATGTC 360  
 DB 301 GTTCCTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGTACAGATTATGTC 360  
 OY 301 GTTCCTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGTACAGATTATGTC 360  
 DB 301 GTTCCTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGTACAGATTATGTC 360  
 OY 361 AACTTGTGAGAGCATTTGGAATGATTTGCTTTTTCAGCTTGACAGAGTTTGGAGT 420  
 DB 361 AACTTGTGAGAGCATTTGGAATGATTTGCTTTTTCAGCTTGACAGAGTTTGGAGT 420  
 OY 421 ATGCAAAAGCATATATTTTGGCAAAAGAAATTAATCTCTCTGAAACATCTAAAGATG 480  
 DB 421 ATGCAAAAGCATATATTTTGGCAAAAGAAATTAATCTCTCTGAAACATCTAAAGATG 480  
 OY 481 AAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTCCCAAAAGACTTCTACAGATG 540  
 DB 481 AAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTCCCAAAAGACTTCTACAGATG 540

OY 541 AACCCGAAATTCCTCTTGCAGAAACAGTCTCAGTGTCCAACTCTTAACCTTGAA 600  
 DB 541 AACCCGAAATTCCTCTTGCAGAAACAGTCTCAGTGTCCAACTCTTAACCTTGAA 600  
 OY 601 CTGTGAGAACTGTGAGACAAAGCAGCGGATATCAACCTCAAAAGAGCTGTCTCATTTG 660  
 DB 601 CTGTGAGAACTGTGAGACAAAGCAGCGGATATCAACCTCAAAAGAGCTGTCTCATTTG 660  
 OY 661 AATTGGATCTGATCTCTTGAAGATACCTTAAATAGGCACTTTTCAGTGTGGAG 720  
 DB 661 AATTGGATCTGATCTCTTGAAGATACCTTAAATAGGCACTTTTCAGTGTGGAG 720  
 OY 721 ATCAGAAATGTGTAACAATCACCCTCAAGAAACAGGAGTAAATCACTTTGATTTG 780  
 DB 721 ATCAGAAATGTGTAACAATCACCCTCAAGAAACAGGAGTAAATCACTTTGATTTG 780  
 OY 781 CAAAAAGCGCTGCTTGTGAATTTTCTGAGACGATATTAACAATTAATCAATCATCAAC 840  
 DB 781 CAAAAAGCGCTGCTTGTGAATTTTCTGAGACGATATTAACAATTAATCAATCATCAAC 840  
 OY 841 CCAATATATGATTTGAACACCACTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 900  
 DB 841 CCAATATATGATTTGAACACCACTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 900  
 OY 901 ATCAGGATGATTTGTTTCAAACTTGCATGTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 960  
 DB 901 ATCAGGATGATTTGTTTCAAACTTGCATGTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 960  
 OY 961 GCTCATATACAGCATGAGAACAGAGGTTTATTTACTCACTAAAGACAGAAATGATGAAA 1020  
 DB 961 GCTCATATACAGCATGAGAACAGAGGTTTATTTACTCACTAAAGACAGAAATGATGAAA 1020  
 OY 1021 AGGCTGATTTCTGATTAATTAAGCAACAGCTTGGCTTACAGAGGCTTACAGAGGCTTACAGAGG 1080  
 DB 1021 AGGCTGATTTCTGATTAATTAAGCAACAGCTTGGCTTACAGAGGCTTACAGAGGCTTACAGAGG 1080  
 OY 1081 GGGCTGGAATTAAGAAACATGTAATGATAGGCGGATCCCAAGACAGAAAGAGTTAG 1140  
 DB 1081 GGGCTGGAATTAAGAAACATGTAATGATAGGCGGATCCCAAGACAGAAAGAGTTAG 1140  
 OY 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGAAATGACGAAATGCTCATGCT 1200  
 DB 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGAAATGACGAAATGCTCATGCT 1200  
 OY 1201 CAGAGATCTTACAGATACGAGATGTTCTTGGATACACTAAATATACAGATTTGAGA 1260  
 DB 1201 CAGAGATCTTACAGATACGAGATGTTCTTGGATACACTAAATATACAGATTTGAGA 1260  
 OY 1261 AAGTTAATGATGTTTCCAGAGATGATGATGTTAGTTCTGTGATGATGATGATGATG 1320  
 DB 1261 AAGTTAATGATGTTTCCAGAGATGATGATGTTAGTTCTGTGATGATGATGATGATGATG 1320  
 OY 1321 GGGAGTCTGATCAATATGCCAAAGTACTGATGATGATGATGATGATGATGATGATG 1380  
 DB 1321 GGGAGTCTGATCAATATGCCAAAGTACTGATGATGATGATGATGATGATGATGATGATG 1380  
 OY 1381 AATATCTGTTCTTACAGAAATAGACTTCTGAGGCTGATGATGATGATGATGATGATGATG 1440  
 DB 1381 AATATCTGTTCTTACAGAAATAGACTTCTGAGGCTGATGATGATGATGATGATGATGATGATG 1440  
 OY 1441 TATGTAAGTGAAGAGTCACTCAATCAATGATGATGATGATGATGATGATGATGATGATG 1500  
 DB 1441 TATGTAAGTGAAGAGTCACTCAATCAATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 OY 1501 TTGGGAAACCTATTCGGAAGAGGCAAGCTCCCAACTTAAGCCATGTAAGTGAATATC 1560  
 DB 1501 TTGGGAAACCTATTCGGAAGAGGCAAGCTCCCAACTTAAGCCATGTAAGTGAATATC 1560  
 OY 1561 TAATTAATGAGGATTTGTTACTGAGGCAACATTAATTAAGAGGCTCCCTCACAATA 1620  
 DB 1561 TAATTAATGAGGATTTGTTACTGAGGCAACATTAATTAAGAGGCTCCCTCACAATA 1620



QY 1621 AATTAAAGCGTAAAGAGACCTATCAGCCCTTCATCTGAGATTTTATCAAGAAAG 1680  
Db 1621 AATTAAAGCGTAAAGAGACCTATCAGCCCTTCATCTGAGATTTTATCAAGAAAG 1680  
QY 1681 CAGATTGGGAGTTCACAAAGACCTCTGAAATGATTAATTCAGGAGACTRACCAAGCGAGC 1740  
Db 1681 CAGATTGGGAGTTCACAAAGACCTCTGAAATGATTAATTCAGGAGACTRACCAAGCGAGC 1740  
QY 1741 AGAATGCTCAAGTATGATTAATTCATAGTGTCTGAGATTAATAACAAAGGTGATT 1800  
Db 1741 AGAATGCTCAAGTATGATTAATTCATAGTGTCTGAGATTAATAACAAAGGTGATT 1800  
QY 1801 CTATTCGAAATGAGAAAAATCTTAACCAATAGAAATCACTCGAAAAAGAAATCTGCTTCA 1860  
Db 1801 CTATTCGAAATGAGAAAAATCTTAACCAATAGAAATCACTCGAAAAAGAAATCTGCTTCA 1860  
QY 1861 AAACGAAAGCTGAACCTTAAGAGCAGTATTAAGCAATATGGAATCTGAAATTAATATCC 1920  
Db 1861 AAACGAAAGCTGAACCTTAAGAGCAGTATTAAGCAATATGGAATCTGAAATTAATATCC 1920  
QY 1921 ACAATTCAAAAGCAGCTTAAGAAAGATAGGCTGAGAGAGAGTCTTCTACAGCATATTC 1980  
Db 1921 ACAATTCAAAAGCAGCTTAAGAAAGATAGGCTGAGAGAGAGTCTTCTACAGCATATTC 1980  
QY 1981 ATGCGCTTGAACCTAGTATGATGAAGAAATCTTAAGCCCACTAATTTGATGTAATTCGAAA 2040  
Db 1981 ATGCGCTTGAACCTAGTATGATGAAGAAATCTTAAGCCCACTAATTTGATGTAATTCGAAA 2040  
QY 2041 TTGATATTTGTTCTAGAGAGTGAAGATTAAGAAAAAAAGTACAACCAATATGCTGCTCA 2100  
Db 2041 TTGATATTTGTTCTAGAGAGTGAAGATTAAGAAAAAAAGTACAACCAATATGCTGCTCA 2100  
QY 2101 GGCACACAGAAACCTTAACCTCATGAGAGTAAAGAACTGCAACTGAGACCAAGAGA 2160  
Db 2101 GGCACACAGAAACCTTAACCTCATGAGAGTAAAGAACTGCAACTGAGACCAAGAGA 2160  
QY 2161 GTAAACAAGCCAAATGAACAGACAAGTAAAGACATGACAGAGTACTTCCAGAGCTGA 2220  
Db 2161 GTAAACAAGCCAAATGAACAGACAAGTAAAGACATGACAGAGTACTTCCAGAGCTGA 2220  
QY 2221 AGTTAAAGATGCACTGCTGTTCTTTTACTAGTGTCAATATACAGTGAATTAAGAAAT 2280  
Db 2221 AGTTAAAGATGCACTGCTGTTCTTTTACTAGTGTCAATATACAGTGAATTAAGAAAT 2280  
QY 2281 TTGTCATCTCAGCTTCCTCAAGAGAAAGAAAAAGAAACTAGAAACAGTAAAGTGT 2340  
Db 2281 TTGTCATCTCAGCTTCCTCAAGAGAAAGAAAAAGAAACTAGAAACAGTAAAGTGT 2340  
QY 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCGAAACTG 2400  
Db 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCGAAACTG 2400  
QY 2401 AAAGATGCTGAGAGAGTATTTCAATTTGCTGACTGATGATTAAGGCACTGAG 2460  
Db 2401 AAAGATGCTGAGAGAGTATTTCAATTTGCTGACTGATGATTAAGGCACTGAG 2460  
QY 2461 AAAGATCTGCTTACTGGAAGTATGCACTCTAGGAGAGGCAAAAACAGAACCAATTAAT 2520  
Db 2461 AAAGATCTGCTTACTGGAAGTATGCACTCTAGGAGAGGCAAAAACAGAACCAATTAAT 2520  
QY 2521 GTGTCAGTCAAGTGTGACAGATTTGAAAAACCCCAAGGAGCTAATTTCAATGTTTCCAAAG 2580  
Db 2521 GTGTCAGTCAAGTGTGACAGATTTGAAAAACCCCAAGGAGCTAATTTCAATGTTTCCAAAG 2580  
QY 2581 ATATATAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGTCC 2640  
Db 2581 ATATATAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGTCC 2640  
QY 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATTTGATGCTCAGTATTTTGGAGAAATCAT 2700  
Db 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATTTGATGCTCAGTATTTTGGAGAAATCAT 2700  
QY 2701 TCAAGTTTCCAAAGCGGCTCATTTGCTCTGTTTCAAAATCCAGAAATTCAGAGAGG 2760

Db 2701 TCAAGTTTCCAAAGCGGCTCATTTGCTCTGTTTCAAAATCCAGAAATTCAGAGAGG 2760  
QY 2761 AATGTGAACATTTCTGCCCCAGCTGTGGGCTTAAAGAAACAAAGTCCAAAGCACTT 2820  
Db 2761 AATGTGAACATTTCTGCCCCAGCTGTGGGCTTAAAGAAACAAAGTCCAAAGCACTT 2820  
QY 2821 TTGAATGTGAACAAAGAGAGAAATCAAGAAAGAAATGAGTCTAATTCAGAGCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAGAGAGCTGTAGGTTTGTCTATCATCTCATGTTTCAGAGGCA 3000  
QY 2881 AGACAGTTAATATCAGCTGACAGGCTTCTGTTGTTGTCAGAAAGATTAAGCCAGTTGATA 2940  
Db 2881 AGACAGTTAATATCAGCTGACAGGCTTCTGTTGTTGTCAGAAAGATTAAGCCAGTTGATA 2940  
QY 2941 ATGCGAAATGTAGTATCAAGAGAGGCTGTAGGTTTGTCTATCATCTCATGTTTCAGAGGCA 3000  
Db 2941 ATGCGAAATGTAGTATCAAGAGAGGCTGTAGGTTTGTCTATCATCTCATGTTTCAGAGGCA 3000  
QY 3001 ACGAACTGAGACTATTACTCCAAATTAACATGAGCTTTTACAAAACCCATATGCTATAC 3060  
Db 3001 ACGAACTGAGACTATTACTCCAAATTAACATGAGCTTTTACAAAACCCATATGCTATAC 3060  
QY 3061 CACCACCTTTTCCATCAGTCAATTTGTTAAACCTAAATGTAAAGAAAATCTGCTAGAGG 3120  
Db 3061 CACCACCTTTTCCATCAGTCAATTTGTTAAACCTAAATGTAAAGAAAATCTGCTAGAGG 3120  
QY 3121 AAAAATTTGAGAGACATTTCAATGTCACCTGAAAGAAAGTGGAAATGAGAAACATTCGAA 3180  
Db 3121 AAAAATTTGAGAGACATTTCAATGTCACCTGAAAGAAAGTGGAAATGAGAAACATTCGAA 3180  
QY 3181 GTACAGTGCAGACATTAAGCCGTATTAACATTAAGAAAGAAATGTTTAAAGAACCCAGCT 3240  
Db 3181 GTACAGTGCAGACATTAAGCCGTATTAACATTAAGAAAGAAATGTTTAAAGAACCCAGCT 3240  
QY 3241 CAAGCAATATTAATGAAGTGAAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3300  
Db 3241 CAAGCAATATTAATGAAGTGAAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3300  
QY 3301 TAGTTCAGAGTGAAGAAACATTCAGAGAGAACTAGTGAAGAAAGAGGCGCAAAATTTGA 3360  
Db 3301 TAGTTCAGAGTGAAGAAACATTCAGAGAGAACTAGTGAAGAAAGAGGCGCAAAATTTGA 3360  
QY 3361 ATGCTATGCTTAGATTTAGGGGTTTTCGAACCTGAGGCTGTATTAACAAAGTCTCTCGAA 3420  
Db 3361 ATGCTATGCTTAGATTTAGGGGTTTTCGAACCTGAGGCTGTATTAACAAAGTCTCTCGAA 3420  
QY 3421 GTAAATTTGATGACATCTGAATTAAGAAAGCAAGATATGAAGAAAGTACTCAGACTGTA 3480  
Db 3421 GTAAATTTGATGACATCTGAATTAAGAAAGCAAGATATGAAGAAAGTACTCAGACTGTA 3480  
QY 3481 ATACAGATTTCTCCATATCTGATTCAGATTAATCTTGAAGACAGCTATGGGAAGTATGTC 3540  
Db 3481 ATACAGATTTCTCCATATCTGATTCAGATTAATCTTGAAGACAGCTATGGGAAGTATGTC 3540  
QY 3541 ATGCATCTCAGGTTTGTCTGACAGACCTGATGACCTGTTAGATGATGCTGAATTAAGG 3600  
Db 3541 ATGCATCTCAGGTTTGTCTGACAGACCTGATGACCTGTTAGATGATGCTGAATTAAGG 3600  
QY 3601 AAGTACTAGTTTTGTGCAAAATGACATTAAGGAAAGTTCGCTGTTTATGCAAAAGCG 3660  
Db 3601 AAGTACTAGTTTTGTGCAAAATGACATTAAGGAAAGTTCGCTGTTTATGCAAAAGCG 3660  
QY 3661 TCCAGAAAGGAGACTTGAAGAGAGTCTTACCCCTTACCCATACACATTTGCTGCTAGG 3720  
Db 3661 TCCAGAAAGGAGACTTGAAGAGAGTCTTACCCCTTACCCATACACATTTGCTGCTAGG 3720  
QY 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGAGTCTCAGAGAAACCTTATCTGATGAGATG 3780  
Db 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGAGTCTCAGAGAAACCTTATCTGATGAGATG 3780  
QY 3781 AAGAGCTTCCCTGCTTCCAAACATTTGTTATTTGTTAAAGTAAACATATATACCTTCTGAGT 3840  
Db 3781 AAGAGCTTCCCTGCTTCCAAACATTTGTTATTTGTTAAAGTAAACATATATACCTTCTGAGT 3840

Dp	3781	AAGACCTTC	CCGCTTCC	AACACACT	GTGTA	TTTGGTA	AAAGTA	AAACAT	TA	ACTTCT	CAGT	3840	
QY	3841	CTACTAG	CGATAG	CACCG	TGTGCT	ACCAG	AGTGCT	CTCTCT	TAAG	AACACAG	AGAGAT	3900	
Dp	3841	CTACTAG	CGATAG	CACCG	TGTGCT	ACCAG	AGTGCT	CTCTCT	TAAG	AACACAG	AGAGAT	3900	
QY	3901	TATCTTA	GAAAGAT	TGCTTAA	TATGCT	CAGTA	AC	CAGT	TA	AT	TGGCA	AGCAT	3960
Dp	3901	TATCTTA	GAAAGAT	TGCTTAA	TATGCT	CAGTA	AC	CAGT	TA	AT	TGGCA	AGCAT	3960
QY	3961	AGGAAC	ATCACC	TTAGT	GAGGA	AAACAA	AT	GTCT	AGCT	TGTTTCT	TCAC	4020	
Dp	3961	AGGAAC	ATCACC	TTAGT	GAGGA	AAACAA	AT	GTCT	AGCT	TGTTTCT	TCAC	4020	
QY	4021	GTGAAT	TGGAG	ACCTT	GACTG	CAAT	TAAAC	ACCC	AGAGAT	CTTCTT	CTGAT	4080	
Dp	4021	GTGAAT	TGGAG	ACCTT	GACTG	CAAT	TAAAC	ACCC	AGAGAT	CTTCTT	CTGAT	4080	
QY	4081	CCAAACA	ATGAG	GGCAT	AGTCTG	CAAC	CCAGG	AGTGGT	CTG	TGAGT	CAC	4140	
Dp	4081	CCAAACA	ATGAG	GGCAT	AGTCTG	CAAC	CCAGG	AGTGGT	CTG	TGAGT	CAC	4140	
QY	4141	TTTCAG	ATGAT	GAGAA	AGAGAC	GGGCTT	GGA	AAAT	TA	TC	AAG	4200	
Dp	4141	TTTCAG	ATGAT	GAGAA	AGAGAC	GGGCTT	GGA	AAAT	TA	TC	AAG	4200	
QY	4201	TGGATT	CAACCT	TAGTG	AAGCAG	ATCTGG	GTG	GAG	AGTGA	AAACA	AGCTCT	4260	
Dp	4201	TGGATT	CAACCT	TAGTG	AAGCAG	ATCTGG	GTG	GAG	AGTGA	AAACA	AGCTCT	4260	
QY	4261	ACTGCT	CAGG	GCAT	CTCTC	TAG	AGTGA	CA	TTTTA	ACCCT	CAG	4320	
Dp	4261	ACTGCT	CAGG	GCAT	CTCTC	TAG	AGTGA	CA	TTTTA	ACCCT	CAG	4320	
QY	4321	AACAT	AACCT	GATTA	AGCTCC	CAGCAG	GAAT	GGCG	TA	CTGA	AGCTG	4380	
Dp	4321	AACAT	AACCT	GATTA	AGCTCC	CAGCAG	GAAT	GGCG	TA	CTGA	AGCTG	4380	
QY	4381	ATGGAG	CCAG	CTTCT	TAAC	AGCTT	AC	CTTCA	TAT	AA	GT	4440	
Dp	4381	ATGGAG	CCAG	CTTCT	TAAC	AGCTT	AC	CTTCA	TAT	AA	GT	4440	
QY	4441	ACCTG	CGAA	ATCC	AGAA	GCAC	ATC	AGAA	AAAG	CGAT	TTA	4500	
Dp	4441	ACCTG	CGAA	ATCC	AGAA	GCAC	ATC	AGAA	AAAG	CGAT	TTA	4500	
QY	4501	GTGAAT	ACCCT	ATATA	AGCC	AGAT	CCAG	AGG	CC	TTTCTG	CTG	4560	
Dp	4501	GTGAAT	ACCCT	ATATA	AGCC	AGAT	CCAG	AGG	CC	TTTCTG	CTG	4560	
QY	4561	CAGAT	AGTTCT	TA	CCAGT	AAATTA	AA	AGAC	AGT	GGAA	AGT	4620	
Dp	4561	CAGAT	AGTTCT	TA	CCAGT	AAATTA	AA	AGAC	AGT	GGAA	AGT	4620	
QY	4621	GGCCAT	CTAT	TAG	ATAT	AGTGT	GCAT	AG	CA	GTGCTG	GGG	4680	
Dp	4621	GGCCAT	CTAT	TAG	ATAT	AGTGT	GCAT	AG	CA	GTGCTG	GGG	4680	
QY	4681	ACTAC	CCAT	CTC	TA	AGAG	AGCT	CA	TTA	AGT	GTG	4740	
Dp	4681	ACTAC	CCAT	CTC	TA	AGAG	AGCT	CA	TTA	AGT	GTG	4740	
QY	4741	AGT	CGGG	CCAC	AGAT	TGAC	GGAA	AC	AT	CTTAC	CTG	4800	
Dp	4741	AGT	CGGG	CCAC	AGAT	TGAC	GGAA	AC	AT	CTTAC	CTG	4800	
QY	4801	CCCCT	TACT	TGAA	TCTG	GAAT	CAG	CC	CTTCT	TG	AT	4860	
Dp	4801	CCCCT	TACT	TGAA	TCTG	GAAT	CAG	CC	CTTCT	TG	AT	4860	
QY	4861	AAGAC	AGAC	CCCC	CAG	AGT	CAG	CT	GTG	TGG	CA	4920	
Dp	4861	AAGAC	AGAC	CCCC	CAG	AGT	CAG	CT	GTG	TGG	CA	4920	

OY		4921	AAGTTCCCAATTGAAGTTCGAGAAATCTCCCCAAGAATCGACGTCTCCTCATACTACTG	4980
Db		4921	AAGTTCCCAATTGAAGTTCGAGAAATCTCCCCAAGAATCGACGTCTCCTCATACTACTG	4980
OY		4981	ATACTGCTGGGATTAATGCACAATGGAAGAAAGTGTTAGCAGAGGAGAACCCAGAAATTGACAG	5040
Db		4981	ATACTGCTGGGATTAATGCACAATGGAAGAAAGTGTTAGCAGAGGAGAACCCAGAAATTGACAG	5040
OY		5041	CTTCAACAGAAAAGGCTCAACAAAGAATGCCATGTTGGCTGTGACCCTGACCCAGAG	5100
Db		5041	CTTCAACAGAAAAGGCTCAACAAAGAATGCCATGTTGGCTGTGACCCTGACCCAGAG	5100
OY		5101	AATTATCTCTGTCGACAAAGTTTGGCAGAAAAACACACATCATCTTAATCTAATTA	5160
Db		5101	AATTATCTCTGTCGACAAAGTTTGGCAGAAAAACACACATCATCTTAATCTAATTA	5160
OY		5161	CTGAGAGACTACTCATGTTGTTATGAAGAACAGATGTCGAGTTGTTGTGTGAACGACAC	5220
Db		5161	CTGAGAGACTACTCATGTTGTTATGAAGAACAGATGTCGAGTTGTTGTGTGAACGACAC	5220
OY		5221	TGAAATATTTCTAGGAATTGGCGGAGAGAAAATGGATAGTTATTTCTGGCTGACCC	5280
Db		5221	TGAAATATTTCTAGGAATTGGCGGAGAGAAAATGGATAGTTATTTCTGGCTGACCC	5280
OY		5281	AGTCATTATTAAGAAAGAAAATGCTGATGATGACATGATTTTGAATCAGTAGAGATGAG	5340
Db		5281	AGTCATTATTAAGAAAGAAAATGCTGATGATGACATGATTTTGAATCAGTAGAGATGAG	5340
OY		5341	TCAATGGAAGAAACCCACCAAGTCCCAAGCAGACAGAGAAATCCCAGAGACAGAAAGATCT	5400
Db		5341	TCAATGGAAGAAACCCACCAAGTCCCAAGCAGAGAGAAATCCCAGAGACAGAAAGATCT	5400
OY		5401	TCAGGGGGCTAGAAATCTGTGTTGCTATGGGCCCTTGACCAATGCCCCACATCAACTGG	5460
Db		5401	TCAGGGGGCTAGAAATCTGTGTTGCTATGGGCCCTTGACCAATGCCCCACATCAACTGG	5460
OY		5461	AATGATGATGTCAGAGTGTGTGTTGCTGTTGATGTAAGAGAGCTTCATCATTCACCCCTTG	5520
Db		5461	AATGATGATGTCAGAGTGTGTGTTGCTGTTGATGTAAGAGAGCTTCATCATTCACCCCTTG	5520
OY		5521	GCACAGGTGTCACCCCAATTTGTTGTTGTCAGCCAGATGTCCTGGACAGAGACAAATGGCT	5580
Db		5521	GCACAGGTGTCACCCCAATTTGTTGTTGTCAGCCAGATGTCCTGGACAGAGACAAATGGCT	5580
OY		5581	TCCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTGACCCGGAGAGTGGTGTGGACA	5640
Db		5581	TCCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTGACCCGGAGAGTGGTGTGGACA	5640
OY		5641	GTTGTAGCACTACACAGTGCAGAGAGAGCTGACACCTACTCATATACCCAGATGCCCCACA	5700
Db		5641	GTTGTAGCACTACACAGTGCAGAGAGAGCTGACACCTACTCATATACCCAGATGCCCCACA	5700
OY		5701	GCCACTACTGA 5711	
Db		5701	GCCACTACTGA 5711	
RESULT 11				
AAV46462..				
ID		AAV46462 standard; cDNA; 5711 BP.		
XX		AAV46462;		
AC				
XX		18-NOV-1998 (first entry)		
DT				
XX		Human BRCA1 omi2 polymorphism #5 cDNA.		
DE				
XX		BRCA1: omi2; human: breast and ovarian cancer predisposing gene:		
KM		polymorphism; susceptibility; anti-oncogene; tumour suppressor;		
KW		chromosome 11q;. ss.		
XX				
OS		Homo sapiens		

XX	Key	Location/Qualifiers
FH	CDS	120..5711
FT		/tag= a
FT		/product= "BRCA1 omi2 protein"
FT	Variation	3667
FT		/*tag= b
FT		/note= "This polymorphic variation can be an A or G nucleotide"
FT		
PN		US5750400-A.
PD		12-MAY-1998.
XX		
PF		12-FEB-1997; 97US-0798691.
PR		12-FEB-1996; 96US-0598591.
PR		12-FEB-1997; 97US-0798691.
PA		(ONCORMED INC.
PI		Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
PI		Scheelter DB, Zeng B;
DR		WPI: 1998-296774/26.
XX		
PT		BRCA1 omi gene coding sequences - useful for distinguishing between
PT		polymorphisms and mutation(s) in the screening for disposition to
PT		breast or ovarian cancer
PS		Claim 2e; Page -: 54pp; English.
XX		
CC		This sequence encodes a human BRCA1 (breast and ovarian cancer
CC		predisposing gene) omi2 gene in which a polymorphic variation occurs at
CC		nucleotide 3667. This sequence and other polymorphic variations of this
CC		sequence are useful for the identification of an individual who may or
CC		may not have an increased susceptibility to breast or ovarian cancer.
CC		The sequences used identify gene changes which are due to polymorphisms,
CC		rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC		suppressor) which is involved in genetic inheritance of cancers,
CC		especially breast and ovarian cancer. It is found at human chromosome
CC		17q which is known to be linked to cancer susceptibility, especially
CC		breast cancer. Cells containing a mutation in this gene lose the
CC		wild-type function of BRCA1 and are more susceptible to cancers.
CC		NOTE: This sequence does not appear in the specification but has been
CC		created from the wild type BRCA1 omi2 gene represented in AAU46449.
XX		
SS		Sequence 5711 BP: 1955 A; 1098 C; 1274 G; 1383 T; 1 other:
Query Match	100.0%; Score 5709; DB 19; Length 5711;	
Best Local Similarity	100.0%; Pred. No.: 0;	
Matches 5709; Conservative . 1; Mismatches 1; Indels 0; Gaps 0;		
OY	1 AGCTCGCTGAGACTTCTTCGGACCCCGCACACAGGCTGTGGGTTTCTCAGATPACTGGGCC	60
DY		
DY	1 AGCTCGCTGAGACTTCTTCGGACCCCGCACACAGGCTGTGGGTTTCTCAGATPACTGGGCC	60
OY	61 CCTGCGCTCAGAGAGCCCTTACCCTCTGCCTGTGGGTAAAGTTCAATTGAACAAGAAA	120
DY		
DY	61 CCTGCGCTCAGAGAGCCCTTACCCTCTGCCTGTGGGTAAAGTTCAATTGAACAAGAAA	120
OY	121 TGGATTTATCTGCTCTTCGCTTGAGAGAATACAAAAATGATTAATTCCTTGCAGAAA	180
DY		
DY	121 TGGATTTATCTGCTCTTCGCTTGAGAGAATACAAAAATGATTAATTCCTTGCAGAAA	180
OY	181 TCATTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACCTGTCTCCACAAAGTGTGACC	240
DY		
DY	181 TCATTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACCTGTCTCCACAAAGTGTGACC	240
OY	241 ACAATATTTTGAATTTTGGATTCGTGTAACCTTCTTCACACAGAAAGAGGCTTCACAGT	300
DY		
DY	241 ACAATATTTTGAATTTTGGATTCGTGTAACCTTCTTCACACAGAAAGAGGCTTCACAGT	300

QY	301	GTCCCTTATGTAAGATGATATATACCAAAAGAGGCTCTCAAGAAAGTACGAGATTTACTC	360
Db	301	GTCCCTTATGTAAGATGATATATATACCAAAAGAGGCTCTCAAGAAAGTACGAGATTTACTC	360
QY	361	AACCTGTGAAGAGCATTTGAAAAATCATTTTGCTTTCAGCTTGACACAGATTGGAGT	420
Db	361	AACCTGTGAAGAGCATTTGAAAAATCATTTTGCTTTCAGCTTGACACAGATTGGAGT	420
QY	421	ATGCAACAGCATATATTTTGCAAAAAAGGAAATTAACCTCTCGTAACATCTAAAAAGATG	480
Db	421	ATGCAACAGCATATATTTTGCAAAAAAGGAAATTAACCTCTCGTAACATCTAAAAAGATG	480
QY	481	AAGTTTCTATCATCCAAAGTATGGGCTACAGAACCGTGCCAAAAAGACTTCTACAGAGT	540
Db	481	AAGTTTCTATCATCCAAAGTATGGGCTACAGAACCGTGCCAAAAAGACTTCTACAGAGT	540
QY	541	AAOCGGAAATTCCTTCCTTGACAGGAACAGCTAGTGTCACATCTCTAACCTTGGA	600
Db	541	AAOCGGAAATTCCTTCCTTGACAGGAACAGCTAGTGTCACATCTCTAACCTTGGA	600
QY	601	CTGTGAGAACTCTGAGCAAAAGCAGCGGATACAACTCTCAAAAGACGCTGCTCATTTG	660
Db	601	CTGTGAGAACTCTGAGCAAAAGCAGCGGATACAACTCTCAAAAGACGCTGCTCATTTG	660
QY	661	AATTGGGATCTGATCTTCTCTGAGATACCGTTAATTAAGGCACATTATTGCACTGGAG	720
Db	661	AATTGGGATCTGATCTTCTCTGAGATACCGTTAATTAAGGCACATTATTGCACTGGAG	720
QY	721	ATCAGAAATTTGTTACAAATCACCCCTCAAGGACCAAGGATGAATTCAGTTGGATTCTG	780
Db	721	ATCAGAAATTTGTTACAAATCACCCCTCAAGGACCAAGGATGAATTCAGTTGGATTCTG	780
QY	781	CAAAAAGGCGCTGTTGAAATTTTCTGAGAGGATGTAACTTAACATCATCTAC	840
Db	781	CAAAAAGGCGCTGTTGAAATTTTCTGAGAGGATGTAACTTAACATCATCTAC	840
QY	841	CCAGTAAATATGATTTGAACACCACTAGGAAGCGTGCACTGAGAGGCATCCAGAAAGT	900
Db	841	CCAGTAAATATGATTTGAACACCACTAGGAAGCGTGCACTGAGAGGCATCCAGAAAGT	900
QY	901	ATCAGGATGATTCCTTTCAAACTGCAATGGAGGCCATGTGGGCAAAATCTCATGGCA	960
Db	901	ATCAGGATGATTCCTTTCAAACTGCAATGGAGGCCATGTGGGCAAAATCTCATGGCA	960
QY	961	GCTCATTTACAGATGAGAACAGCACTGAAAGCGCTGATACCTAAAGACAGATGATAGAA	1020
Db	961	GCTCATTTACAGATGAGAACAGCACTGAAAGCGCTGATACCTAAAGACAGATGATAGAA	1020
QY	1021	AGCGTGAATTCGTATATTAANAGCAACAGCGCTGTAGCAAGGAGCCACATPAACAGAT	1080
Db	1021	AGCGTGAATTCGTATATTAANAGCAACAGCGCTGTAGCAAGGAGCCACATPAACAGAT	1080
QY	1081	GGCGTGAAGTAAGGAACATGTATGTATAGCGGACCTCCAGCAACAAAAAAGGTAG	1140
Db	1081	GGCGTGAAGTAAGGAACATGTATGTATAGCGGACCTCCAGCAACAAAAAAGGTAG	1140
QY	1141	ATCTGAAATGCTGATCCCTGTGTGAGAAAGAAATGGAATPAAGCAGAACTGCCATCT	1200
Db	1141	ATCTGAAATGCTGATCCCTGTGTGAGAAAGAAATGGAATPAAGCAGAACTGCCATCT	1200
QY	1201	CAGAGAAATCTAGAGATCTGTAAGATGTTCTTGATTAACACTTAATAGCAGCATTCGA	1260
Db	1201	CAGAGAAATCTAGAGATCTGTAAGATGTTCTTGATTAACACTTAATAGCAGCATTCGA	1260
QY	1261	AAGTTAATGATGGTTTTCCAGAAAGTATGAACCTTTAGTTCTGATGACACATGATG	1320
Db	1261	AAGTTAATGATGGTTTTCCAGAAAGTATGAACCTTTAGTTCTGATGACACATGATG	1320
QY	1321	GGGAGTCTGATCAAAATCCCAAAGTACGTATGATTTAGACGTTCTTAATAGGTAGATG	1380
Db	1321	GGGAGTCTGATCAAAATCCCAAAGTACGTATGATTTAGACGTTCTTAATAGGTAGATG	1380
QY	1381	AATATTCGTCTTCAGAGAAATAGACTTACTGCGCAGTATCTCATAGGCTTTAA	1440

```
1381 AATATCTGCTCTTCAGAGAAATAGACTTACGGCAGCATCTCATGAGCTTAA 1440
1441 TATGTAAAGTGAAGGCTCACCACCAATCAGTAGAGAAATATGGAAGCAAAATAT 1500
1441 TATGTAAAGTGAAGGCTCACCACCAATCAGTAGAGAAATATGGAAGCAAAATAT 1500
1501 TTGGGAAAACCTTTCGGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAAGTGAATC 1560
1501 TTGGGAAAACCTTTCGGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAAGTGAATC 1560
1561 TAAATTATAGAGCAATTTGTTACTGAGCCACAGATATATACAGAGCGTCCCTCAAA 1620
1561 TAAATTATAGAGCAATTTGTTACTGAGCCACAGATATATACAGAGCGTCCCTCAAA 1620
1621 AATTAAGGCTAAAGAGCACTACATCAGGCTTCCTCCCTGAGATTTTATCAAGAA 1680
1621 AATTAAGGCTAAAGAGCACTACATCAGGCTTCCTCCCTGAGATTTTATCAAGAA 1680
1681 CAGATTTGGCAGTTTCAAAAAGACTCTGAAATGATTAATACAGGAACTTACCAAG 1740
1681 CAGATTTGGCAGTTTCAAAAAGACTCTGAAATGATTAATACAGGAACTTACCAAG 1740
1741 AGATGCTCAGATGATGATATTTACTATATAGTGTCTAGAGATTAACAAAGGTGAT 1800
1741 AGATGCTCAGATGATGATATTTACTATATAGTGTCTAGAGATTAACAAAGGTGAT 1800
1801 CTATTCAGATGAGAAAATCTTACCCATAGAAATACAGGAAAGAAATCTGCTTCA 1860
1801 CTATTCAGATGAGAAAATCTTACCCATAGAAATACAGGAAAGAAATCTGCTTCA 1860
1861 AAAGCAAGCTGAGCTATATAGCAGAGATTAAGCAATATGAACTGAACTTAAATAT 1920
1861 AAAGCAAGCTGAGCTATATAGCAGAGATTAAGCAATATGAACTGAACTTAAATAT 1920
1921 ACATTTCAAAAGCACTTAAAGAAATAGGCTGAGAGGAAAGTCTTCTACCAAGCAT 1980
1921 ACATTTCAAAAGCACTTAAAGAAATAGGCTGAGAGGAAAGTCTTCTACCAAGCAT 1980
1981 ATGCGCTTGAACAGTATAGTATAGAAATTAAGCCACCTAATTTAGTGAATTTGAAA 2040
1981 ATGCGCTTGAACAGTATAGTATAGAAATTAAGCCACCTAATTTAGTGAATTTGAAA 2040
2041 TTGATAGTGTCTTACAGTGAAGAGATTAAGAAAAGAAAGTACAAACCAATGCG 2100
2041 TTGATAGTGTCTTACAGTGAAGAGATTAAGAAAAGAAAGTACAAACCAATGCG 2100
2101 GGCACAGCAGAAACCTTACAACTCATGAGAGTAAAGAACTGCAACTGGAGCCAA 2160
2101 GGCACAGCAGAAACCTTACAACTCATGAGAGTAAAGAACTGCAACTGGAGCCAA 2160
2161 GTTAAAGCCAAATGAGACAGACAAAGTAAAGACATGAGTACTTTCCAGAGTGA 2220
2161 GTTAAAGCCAAATGAGACAGACAAAGTAAAGACATGAGTACTTTCCAGAGTGA 2220
2221 AGTTAAACAATGCACTGCTTCTTACTAAGTGTTCAAATACAGAGTGAATTAAG 2280
2221 AGTTAAACAATGCACTGCTTCTTACTAAGTGTTCAAATACAGAGTGAATTAAG 2280
2281 TTGTCAATCTGACCTTCCAGAGAGAAAGAAAGAAAGTAAAGAGTAAAGTGT 2340
2281 TTGTCAATCTGACCTTCCAGAGAGAAAGAAAGAAAGTAAAGAGTAAAGTGT 2340
2281 TTGTCAATCTGACCTTCCAGAGAGAAAGAAAGAAAGTAAAGAGTAAAGTGT 2340
2281 TTGTCAATCTGACCTTCCAGAGAGAAAGAAAGAAAGTAAAGAGTAAAGTGT 2340
2341 GTTAAATGCTGAGAGCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAAC 2400
2341 GTTAAATGCTGAGAGCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAAC 2400
2401 AAAGATCTGTAGAGAGTATTTCAATTTGTTAGTGTACTGATTTAGGCACTGAG 2460
2401 AAAGATCTGTAGAGAGTATTTCAATTTGTTAGTGTACTGATTTAGGCACTGAG 2460
2461 AAAGATCTGTAGAGAGTATTTCAATTTGTTAGTGTACTGATTTAGGCACTGAG 2520
2461 AAAGATCTGTAGAGAGTATTTCAATTTGTTAGTGTACTGATTTAGGCACTGAG 2520
```

---

```
D 2461 AAAGATCTGTAGAGAGTATTTCAATTTGTTAGTGTACTGATTTAGGCACTGAG 2520
Q 2521 GTGTGAGTCAAGTGTGAGCACTTTGAAAACCCCAAGGACATTTCAATGTTGTTCAAG 2580
D 2521 GTGTGAGTCAAGTGTGAGCACTTTGAAAACCCCAAGGACATTTCAATGTTGTTCAAG 2580
Q 2581 ATATATAGAAATGACACAGAGAGGCTTAAATATGATTCATTTGGACATGAACTAAC 2640
D 2581 ATATATAGAAATGACACAGAGAGGCTTAAATATGATTCATTTGGACATGAACTAAC 2640
Q 2641 GGGAAACACATATGAAATGAAAGAAAGTAACTTGTAGTCTGATTTTTCAGATAC 2700
D 2641 GGGAAACACATATGAAATGAAAGAAAGTAACTTGTAGTCTGATTTTTCAGATAC 2700
Q 2701 TCAAGGTTTCAAGGCGGACATCTTGTGCTGTTTCAATCCAGAAATTCAGAAAGAG 2760
D 2701 TCAAGGTTTCAAGGCGGACATCTTGTGCTGTTTCAATCCAGAAATTCAGAAAGAG 2760
Q 2761 AATGTCACATCTCTGCGCCTGCTGCTTAAAGAAACAAAGTCCAAAGTCACTT 2820
D 2761 AATGTCACATCTCTGCGCCTGCTGCTTAAAGAAACAAAGTCCAAAGTCACTT 2820
Q 2821 TTGATGTTGACAAAGAGAAATCAGAGAAAGATGATGATTAATTCACAGCTGTAC 2880
D 2821 TTGATGTTGACAAAGAGAAATCAGAGAAAGATGATGATTAATTCACAGCTGTAC 2880
Q 2881 AGACAGTTATATCAGTGCAGGCTTCTGCTGTTGCTGAGAAAGTAAAGCCAGTTGATA 2940
D 2881 AGACAGTTATATCAGTGCAGGCTTCTGCTGTTGCTGAGAAAGTAAAGCCAGTTGATA 2940
Q 2941 ATGCAAAATGATATCAAGAGAGGCTTGTGCTTATCATTCTCAGTTCAAGGCA 3000
D 2941 ATGCAAAATGATATCAAGAGAGGCTTGTGCTTATCATTCTCAGTTCAAGGCA 3000
Q 3001 ACAGAACTGACATCTTACCTCCAAATTAACATGCACTTTTACAAACCATATCTATAC 3060
D 3001 ACAGAACTGACATCTTACCTCCAAATTAACATGCACTTTTACAAACCATATCTATAC 3060
Q 3061 CACCACCTTTTCCCATCAAGTCAATTTGTTAAACTTAAATGAAAGAAATCTGTAGAG 3120
D 3061 CACCACCTTTTCCCATCAAGTCAATTTGTTAAACTTAAATGAAAGAAATCTGTAGAG 3120
Q 3121 AAAAATTTGAGAACATTTCAATGTCACTGAAAGAGAAATGGGAAATGGAACATTTCCA 3180
D 3121 AAAAATTTGAGAACATTTCAATGTCACTGAAAGAGAAATGGGAAATGGAACATTTCCA 3180
Q 3181 GTACAGTGAACACAAATTTAGCCGTTAATTAACATTAAGAAATGTTTTTAAAGAGCCACT 3240
D 3181 GTACAGTGAACACAAATTTAGCCGTTAATTAACATTAAGAAATGTTTTTAAAGAGCCACT 3240
Q 3241 CAAGCAATTTAATGAAGTGTGCTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3300
D 3241 CAAGCAATTTAATGAAGTGTGCTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3300
Q 3301 TAGGTTCCAGTGTGAAACATTTCAAGCAGAACTAGTGAAGAAACAGAGGCGCAAAAT 3360
D 3301 TAGGTTCCAGTGTGAAACATTTCAAGCAGAACTAGTGAAGAAACAGAGGCGCAAAAT 3360
Q 3361 ATGCTATGCTTAAATTAAGGGGTTTTCGCAACCTGAGAGTCTATTAACAAAGTCTCTG 3420
D 3361 ATGCTATGCTTAAATTAAGGGGTTTTCGCAACCTGAGAGTCTATTAACAAAGTCTCTG 3420
Q 3421 GTTAAATGTAAGCATCTGAAATTAAGAAAGCAGAAATATGAAGAGTTCAGACTGTTA 3480
D 3421 GTTAAATGTAAGCATCTGAAATTAAGAAAGCAGAAATATGAAGAGTTCAGACTGTTA 3480
Q 3481 ATACAGATTTCTCCATATCTGATTTCAATTAAGTAAAGCAGCTTATGGAAGAGTAC 3540
D 3481 ATACAGATTTCTCCATATCTGATTTCAATTAAGTAAAGCAGCTTATGGAAGAGTAC 3540
Q 3541 ATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAAGATGATGATGAATTAAG 3600
D 3541 ATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAAGATGATGATGAATTAAG 3600
```

OY 3601 AACATCTAGTTTGGTGAATAATGACATTAGGAAGTTCTGCTGTTTATAGCAAAAGCG 3660  
|||||  
Db 3601 AAGATCTAGTTTGGTGAATAATGACATTAGGAAGTTCTGCTGTTTATAGCAAAAGCG 3660  
OY 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTTCACCCATACATATTTGGCTCAG 3720  
|||||  
Db 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTTCACCCATACATATTTGGCTCAG 3720  
OY 3721 GTTACGGAAGAGGGGCCAAGAAATTAGAGTCTCCAGAAAGAACTTATCTAGGAGATG 3780  
|||||  
Db 3721 GTTACGGAAGAGGGGCCAAGAAATTAGAGTCTCCAGAAAGAACTTATCTAGGAGATG 3780  
OY 3781 AAGAGCTTCCCTGCTCCAACTGTTATTTGGTAAAGTAAACAATATACCTTCAG 3840  
|||||  
Db 3781 AAGAGCTTCCCTGCTCCAACTGTTATTTGGTAAAGTAAACAATATACCTTCAG 3840  
OY 3841 CTACTAGGCAATGACACCGTTGCTACCGAGTGTCTGTAGAAACACAGAGAGAAATTTAT 3900  
|||||  
Db 3841 CTACTAGGCAATGACACCGTTGCTACCGAGTGTCTGTAGAAACACAGAGAGAAATTTAT 3900  
OY 3901 TATCATTTGAAGATTAATTAATGATGAGTACAGGTAATATTTGGCAAAAGGCAATCTC 3960  
|||||  
Db 3901 TATCATTTGAAGATTAATTAATGATGAGTACAGGTAATATTTGGCAAAAGGCAATCTC 3960  
OY 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCAGATGGA 4020  
|||||  
Db 3961 AGGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCAGATGGA 4020  
OY 4021 GTGAATTTGAAGACTGACTGCAAAATACAAACACCCAGATCCTTCTTGATTTGTTCTT 4080  
|||||  
Db 4021 GTGAATTTGAAGACTGACTGCAAAATACAAACACCCAGATCCTTCTTGATTTGTTCTT 4080  
OY 4081 CCAAAACAAATGAGGCACTCACTGGAAGCCAGGAGTTGCTGCACTGACAGGAATTTGG 4140  
|||||  
Db 4081 CCAAAACAAATGAGGCACTCACTGGAAGCCAGGAGTTGCTGCACTGACAGGAATTTGG 4140  
OY 4141 TTTCAGATGATGAAGAAAGAGAAAGGCTTGAAGAAATATATCAAGAAAGACAAAGCA 4200  
|||||  
Db 4141 TTTCAGATGATGAAGAAAGAGAAAGGCTTGAAGAAATATATCAAGAAAGACAAAGCA 4200  
OY 4201 TGGATTTCAAACCTTAGTGAAGCAGCATCTGGGTGTAGAGTGAAGCAAGGCTCTGAG 4260  
|||||  
Db 4201 TGGATTTCAAACCTTAGTGAAGCAGCATCTGGGTGTAGAGTGAAGCAAGGCTCTGAG 4260  
OY 4261 ACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATGC 4320  
|||||  
Db 4261 ACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATGC 4320  
OY 4321 AACATTAACCTGATTAAGCTCCAGAGGAATGGCTGAAGTGAAGCTGTTGAGACAGC 4380  
|||||  
Db 4321 AACATTAACCTGATTAAGCTCCAGAGGAATGGCTGAAGTGAAGCTGTTGAGACAGC 4380  
OY 4381 ATGGAGCAGAGCTTCTTAACAGCTACCTTCATTAAGTGACTCTTCTGCCCTTGAG 4440  
|||||  
Db 4381 ATGGAGCAGAGCTTCTTAACAGCTACCTTCATTAAGTGACTCTTCTGCCCTTGAG 4440  
OY 4441 ACCTGGAATCCAGAAACAAAGCAGATCAGAAAAACAGATTTAACTTACAGAAAAAGTA 4500  
|||||  
Db 4441 ACCTGGAATCCAGAAACAAAGCAGATCAGAAAAACAGATTTAACTTACAGAAAAAGTA 4500  
OY 4501 GTGAATTAACCTTATTAACAGAAATCCAGAGGCTTCTGCTGCAAGTTTGAAGTCTG 4560  
|||||  
Db 4501 GTGAATTAACCTTATTAACAGAAATCCAGAGGCTTCTGCTGCAAGTTTGAAGTCTG 4560  
OY 4561 CAGATAGTTCTACCACTAAATAAAGAACACAGAGTGAAGAGTCAATCCCTTCTTAAT 4620  
|||||  
Db 4561 CAGATAGTTCTACCACTAAATAAAGAACACAGAGTGAAGAGTCAATCCCTTCTTAAT 4620  
OY 4621 GCCCATCATTAAGTATAGTGTGTACATGACAGTGTCTGTGGAGTCTTCAAGAAATAGAA 4680  
|||||  
Db 4621 GCCCATCATTAAGTATAGTGTGTACATGACAGTGTCTGTGGAGTCTTCAAGAAATAGAA 4680

OY 4681 ACTACCATCTCAAGAGAGCTCATTAAGTTTGTGATGTGAGAGCAACACTGGAG 4740  
|||||  
Db 4681 ACTACCATCTCAAGAGAGCTCATTAAGTTTGTGATGTGAGAGCAACACTGGAG 4740  
OY 4741 AGTCTGGGCCACACGATTTTGACGAAACATCTTACTTGGCCAAAGCAGATCTAGAGGAA 4800  
|||||  
Db 4741 AGTCTGGGCCACACGATTTTGACGAAACATCTTACTTGGCCAAAGCAGATCTAGAGGAA 4800  
OY 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTCTCTGTATGATGCCCTGAAATCTGATCTCTG 4860  
|||||  
Db 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTCTCTGTATGATGCCCTGAAATCTGATCTCTG 4860  
OY 4861 AAGACAGAGCCCAAGATCAGCTGCTGTTGGCAACATATACATCTTAACTCTGCAATGA 4920  
|||||  
Db 4861 AAGACAGAGCCCAAGATCAGCTGCTGTTGGCAACATATACATCTTAACTCTGCAATGA 4920  
OY 4921 AAGTCCCAATTTGAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTACTG 4980  
|||||  
Db 4921 AAGTCCCAATTTGAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTACTG 4980  
OY 4981 ATACTGCTGGGTATTAATGCAATGGAAGAGTGTGAGCAGAGGAGAAAGCCAAATTTGAC 5040  
|||||  
Db 4981 ATACTGCTGGGTATTAATGCAATGGAAGAGTGTGAGCAGAGGAGAAAGCCAAATTTGAC 5040  
OY 5041 CTTCACAGAAAGGCTCAACAAAGATGTCATGGTGTGCTGGCCCTGACCCAGAG 5100  
|||||  
Db 5041 CTTCACAGAAAGGCTCAACAAAGATGTCATGGTGTGCTGGCCCTGACCCAGAG 5100  
OY 5101 AATTTATGCTCGTGTACAAAGTTTGCCAGAAACACACATCACTTTAACTAATCTAATTA 5160  
|||||  
Db 5101 AATTTATGCTCGTGTACAAAGTTTGCCAGAAACACACATCACTTTAACTAATCTAATTA 5160  
OY 5161 CTGAAGAGACTACTCATGTTGTTATGAAAGACATGCTGAGTTGTGTGGAACGAGAC 5220  
|||||  
Db 5161 CTGAAGAGACTACTCATGTTGTTATGAAAGACATGCTGAGTTGTGTGGAACGAGAC 5220  
OY 5221 TGAATATTTTCTAGGAATTTGCGGAGGAAATGGGTAGTACTATTTCTGGGTGACCC 5280  
|||||  
Db 5221 TGAATATTTTCTAGGAATTTGCGGAGGAAATGGGTAGTACTATTTCTGGGTGACCC 5280  
OY 5281 AGCTATTTAAAGAAAGAAATCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTGG 5340  
|||||  
Db 5281 AGCTATTTAAAGAAAGAAATCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTGG 5340  
OY 5341 TCAATGGAAGAAACCAAGGTCGCAAGAGGAGCAAGAGAAATGCCAGAGAGAGAAATCT 5400  
|||||  
Db 5341 TCAATGGAAGAAACCAAGGTCGCAAGAGGAGCAAGAGAAATGCCAGAGAGAGAAATCT 5400  
OY 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAACATGAGCCACAGATCAACTGG 5460  
|||||  
Db 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAACATGAGCCACAGATCAACTGG 5460  
OY 5461 AATGATGCTACAGCTGTGTGTCTCTGTGTGTAAGAGCTTTCACTTCACTTCACTTGG 5520  
|||||  
Db 5461 AATGATGCTACAGCTGTGTGTCTCTGTGTGTAAGAGCTTTCACTTCACTTCACTTGG 5520  
OY 5521 GCACAGGTGCCCAATTTGTGTGTGTGAGCCAGCAATGCTGTGAGAGAGAGAAATGGCT 5580  
|||||  
Db 5521 GCACAGGTGCCCAATTTGTGTGTGTGAGCCAGCAATGCTGTGAGAGAGAGAAATGGCT 5580  
OY 5581 TCCATGCAATTTGGGAGATGTGAGAGCACTTGTGTGAGCCAGAGAGTGGTGTGGACA 5640  
|||||  
Db 5581 TCCATGCAATTTGGGAGATGTGAGAGCACTTGTGTGAGCCAGAGAGTGGTGTGGACA 5640  
OY 5641 GTGTAGCACTTACCAAGTCCAGAGCTTGACACCTGATATACCCAGATCCCCACA 5700  
|||||  
Db 5641 GTGTAGCACTTACCAAGTCCAGAGCTTGACACCTGATATACCCAGATCCCCACA 5700  
OY 5701 GCCACTACTGA 5711  
|||||  
Db 5701 GCCACTACTGA 5711

RESULT 12
AAT84840
ID AAT84840 standard; cDNA to mRNA; 5712 BP.
XX
AC AAT84840;
XX
DT 21-FEB-1998 (first entry)
XX
DE Human breast and ovarian cancer susceptibility gene BRCA1.
XX
KW BRCA1; breast cancer; ovarian cancer; human;
RN tumour suppressor gene; gene therapy; receptor; ds.
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 120..5711
FT CDS
FT /tag=a
FT /transl_except= (pos:1482..1484, aa:asp)
FT /transl_except= (pos:1581..1583, aa:ser)
FT /transl_except= (pos:1699..1901, aa:gln)
FT /transl_except= (pos:1902..1904, aa:leu)
FT /transl_except= (pos:1905..1907, aa:gln)
FT /transl_except= (pos:1908..1910, aa:leu)
FT /transl_except= (pos:1911..1913, aa:asn)
FT /transl_except= (pos:1914..1916, aa:ile)
FT /transl_except= (pos:1917..1919, aa:met)
FT /transl_except= (pos:4476..4478, aa:val)
FT /transl_except= (pos:4476..4478, aa:val)
FT /transl_except= (pos:4479..4481, aa:leu)
FT /transl_except= (pos:4482..4484, aa:gln)
FT /transl_except= (pos:4689..4691, aa:pro)
XX
PN W09730108-AI.
XX
PD 21-AUG-1997.
XX
PE 19-FEB-1997; 97WO-US03340.
XX
PR 20-FEB-1996; 96US-0603753.
XX
(UYVA-) UNIV VANDERBILT.
PA (UNIV ) UNIV WASHINGTON.
PI
PI Robinson-Benton CL, Szabo CI, Thompson ME;
DR WPI: 1997-434733/40.
DR P-PSDB: AAM23286.
XX
BRCA1 and BRCA2 tumour suppressor gene products - useful to inhibit breast and ovarian cancer cell growth and tumorigenesis, or treat gene linked hereditary or sporadic ovarian or breast cancer
PS Claim 13; Page 54-63; 148bp; English.
XX
This sequence comprises a full-length BRCA1 cDNA. Genetic analysis of familial and ovarian cancer indicates that BRCA1 is a tumour suppressor gene. It encodes a 190 kDa protein (see AAM23286) that is an inhibitor of the growth and proliferation of human breast and ovarian cancer cells. DNA encoding the BRCA1 protein can thus be used in gene therapy methods for the treatment of breast and ovarian cancers. A purified BRCA1 protein can also be used to treat these cancers and, since it is secreted, can be used to identify the BRCA1 receptor and hence to identify BRCA1 protein-mimetic agents which act on the receptor for use in breast and ovarian cancer treatment. The BRCA2 gene (see AAT84841) and BRCA2 protein (see AAM23287) have also been characterised. Methods are claimed for the isolation of BRCA1 or BRCA2 receptors, and for treating or preventing breast and (gene-linked hereditary or sporadic) ovarian cancers using BRCA1 and BRCA2 proteins and genes.
Sequence 5712 BP; 1956 A; 1099 G; 1274 G; 1383 T; 0 other;

Query Match	99.9%;	Score 5707.8;	DB 18;	Length 5712;
Best Local Similarity	100.0%;	Pred. 10.0;		
Matches 5709;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	AGTCGCTGAGACCTCTCTGAGACCCGCGACACAGGCTGTGGGGTTTCTCAGATACCTGGCC	60	
Db	1	AGCTCCCTGAGACTTCTCTGAGACCCGCGACACAGGCTGTGGGGTTTCTCAGATACCTGGCC	60	
QY	61	CTGGCGCTGAGAGGCGCTTCAACCCCTGCTGGGTAACTTCATTGGAACGAAAGAA	120	
Db	61	CTGGCGCTGAGAGGCGCTTCAACCCCTGCTGGGTAACTTCATTGGAACGAAAGAA	120	
QY	121	TGATTTATCTGCTCTCTGCGCTTGAAGAAGTCAAAATGTCATTAAATGCTATGAGAAA	180	
Db	121	TGATTTATCTGCTCTCTGCGCTTGAAGAAGTCAAAATGTCATTAAATGCTATGAGAAA	180	
QY	181	TCTTAAGTGTCCCATCTGTCTGGAGTGTATCAAGAAACCTGTCTCACAAAGTGTGAC	240	
Db	181	TCTTAAGTGTCCCATCTGTCTGGAGTGTATCAAGAAACCTGTCTCACAAAGTGTGAC	240	
QY	241	ACATATTTTGGCAAAATTTTGGCATGCGAAGTCTCAACCCAGAAAGAAAGGCGCTTCACAT	300	
Db	241	ACATATTTTGGCAAAATTTTGGCATGCGAAGTCTCAACCCAGAAAGAAAGGCGCTTCACAT	300	
QY	301	GTCCTTTATGTAAAGTATATACCAAGAGGCTTACAGAAAGTACGACATTTAGTC	360	
Db	301	GTCCTTTATGTAAAGTATATACCAAGAGGCTTACAGAAAGTACGACATTTAGTC	360	
QY	361	AACCTTTGAAGCATATGAAATCATTTGTGTTTTCAGTCTTGACACAGGTTTGGAGT	420	
Db	361	AACCTTTGAAGCATATGAAATCATTTGTGTTTTCAGTCTTGACACAGGTTTGGAGT	420	
QY	421	ATGCAAAACGCTATATATTTTGGCAAAAGAAATTAATCTGCTGAAACATCTAAAGATG	480	
Db	421	ATGCAAAACGCTATATATTTTGGCAAAAGAAATTAATCTGCTGAAACATCTAAAGATG	480	
QY	481	AATTTCTATCTCCAAAGTATGGGCTACAGAAACCGTGGCAAAAGACTTCTACAGATG	540	
Db	481	AATTTCTATCTCCAAAGTATGGGCTACAGAAACCGTGGCAAAAGACTTCTACAGATG	540	
QY	541	AACCCGAAATCCTTCTGTGAGGAAACACAGTCTCACTGCAACTCTTAACCTTGGAA	600	
Db	541	AACCCGAAATCCTTCTGTGAGGAAACACAGTCTCACTGCAACTCTTAACCTTGGAA	600	
QY	601	CTGTGGAACCTGTGAGGACCAAGCAGGGATACACTCAAAAGAGCTGTCTACATTTG	660	
Db	601	CTGTGGAACCTGTGAGGACCAAGCAGGGATACACTCAAAAGAGCTGTCTACATTTG	660	
QY	661	AATTTGATCTATTTCTTCTGAAGATCCGTTATTAAGCAACTTATTCATGTGGGAG	720	
Db	661	AATTTGATCTATTTCTTCTGAAGATCCGTTATTAAGCAACTTATTCATGTGGGAG	720	
QY	721	ATCAGAAATTTGTACAAATCACCCCTCAAGGACACAGGATGAAATCATTTGGATCTG	780	
Db	721	ATCAGAAATTTGTACAAATCACCCCTCAAGGACACAGGATGAAATCATTTGGATCTG	780	
QY	781	CAAAAAAGGCGCTGTGAATTTCTGAGGCGGATGTAAATTAATCTGAACTCATCTAC	840	
Db	781	CAAAAAAGGCGCTGTGAATTTCTGAGGCGGATGTAAATTAATCTGAACTCATCTAC	840	
QY	841	CCAGTAATTAATGATTGGAACACCACTAGAGAAGGTGAGCTGAGAGGCTCCAGAAAGT	900	
Db	841	CCAGTAATTAATGATTGGAACACCACTAGAGAAGGTGAGCTGAGAGGCTCCAGAAAGT	900	
QY	901	ATCAGAGTATGATTGTAACCACTGATGTGAGGCAATGTGCAACAAATTAATCTATGCCA	960	
Db	901	ATCAGAGTATGATTGTAACCACTGATGTGAGGCAATGTGCAACAAATTAATCTATGCCA	960	
QY	961	GCTCATTAAGCATAGAGCAAGCAAGTTTATTAATCACTAAAGCAATAAGATGTAGAAA	1020	
Db	961	GCTCATTAAGCATAGAGCAAGCAAGTTTATTAATCACTAAAGCAATAAGATGTAGAAA	1020	

QY 1021 AGCGTGAATTCCTGAATATAAAGCAAGACGCTGGCTTAGAGAGAGCCAACTAATACAGAT 1080  
DB 1021 AGCGTGAATTCCTGAATATAAAGCAAGACGCTGGCTTAGAGAGAGCCAACTAATACAGAT 1080  
QY 1081 GGGGTGGAAGTAAAGAAACATGTATATAGCGGACTCCGAGACAGAAAAAGGTAG 1140  
DB 1081 GGGGTGGAAGTAAAGAAACATGTATATAGCGGACTCCGAGACAGAAAAAGGTAG 1140  
QY 1141 ATCGAATGCTGAATCCCTGGTGGAGAAAAAGATGGAATTAAGCAAACTGGCATGCT 1200  
DB 1141 ATCGAATGCTGAATCCCTGGTGGAGAAAAAGATGGAATTAAGCAAACTGGCATGCT 1200  
QY 1201 CAGAGAAATCTAGAGATACAGAGATGCTCTGGATTAAGCACTAATAGAGCATTTGAGA 1260  
DB 1201 CAGAGAAATCTAGAGATACAGAGATGCTCTGGATTAAGCACTAATAGAGCATTTGAGA 1260  
QY 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTGAATGCTTTAGCTTGTGATGACTCAGATGATG 1320  
DB 1261 AAGTTAATGAGTGGTTTTCCAGAAAGTGAATGCTTTAGCTTGTGATGACTCAGATGATG 1320  
QY 1321 GGGAGTCTGAATCCAAATGGCCAAATAGCTGATGATGAGACGCTCTAATAGAGTAGATG 1380  
DB 1321 GGGAGTCTGAATCCAAATGGCCAAATAGCTGATGATGAGACGCTCTAATAGAGTAGATG 1380  
QY 1381 AATATCTGCTTCTCAGAGAAAAATAGACTTACTGGCCAGTATCTCATAGAGCTTTAA 1440  
DB 1381 AATATCTGCTTCTCAGAGAAAAATAGACTTACTGGCCAGTATCTCATAGAGCTTTAA 1440  
QY 1441 TATGTAAAAGTGAAGAGGTTCTCCAAATCAGTAGAGAGTAAATTTGAAGACAAATAT 1500  
DB 1441 TATGTAAAAGTGAAGAGGTTCTCCAAATCAGTAGAGAGTAAATTTGAAGACAAATAT 1500  
QY 1501 TTGGGAAACCTATCGAGAGAGGCAAGCCCTCCCACTTAAGCCATGATTAAGTGAATC 1560  
DB 1501 TTGGGAAACCTATCGAGAGAGGCAAGCCCTCCCACTTAAGCCATGATTAAGTGAATC 1560  
QY 1561 TAAATATAGAGAGATTTGTTACTGAGCCAGATTAATCAAGAGCGTCCCTCACAATA 1620  
DB 1561 TAAATATAGAGAGATTTGTTACTGAGCCAGATTAATCAAGAGCGTCCCTCACAATA 1620  
QY 1621 AATTAAGGCTAAAGAGAGACCTAATCAGGCTTCTCATCTGAGGATTTTATCAAGAA 1680  
DB 1621 AATTAAGGCTAAAGAGAGACCTAATCAGGCTTCTCATCTGAGGATTTTATCAAGAA 1680  
QY 1681 CAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAAGGAGCACTAACCAAGGAGC 1740  
DB 1681 CAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAAGGAGCACTAACCAAGGAGC 1740  
QY 1741 AGAATGCTCAAGTGAATATTACTAATAGTGGTCAATGAGAAATTAACCAAGGCTGAT 1800  
DB 1741 AGAATGCTCAAGTGAATATTACTAATAGTGGTCAATGAGAAATTAACCAAGGCTGAT 1800  
QY 1801 CTATATGAGATGAGAAAAATCCTAACCATAAGATACCTCGAAAAAGATCTGCTTTCA 1860  
DB 1801 CTATATGAGATGAGAAAAATCCTAACCATAAGATACCTCGAAAAAGATCTGCTTTCA 1860  
QY 1861 AAAGGAAAGCTGAACCTATAGAGAGAGCTATAGCAATATGAACTGGAATTAATATTC 1920  
DB 1861 AAAGGAAAGCTGAACCTATAGAGAGAGCTATAGCAATATGAACTGGAATTAATATTC 1920  
QY 1921 ACAATTCAAAAGACCTAATAAAGAAATAGGCTGAGAGAGAACTTTCACAGGCAATTC 1980  
DB 1921 ACAATTCAAAAGACCTAATAAAGAAATAGGCTGAGAGAGAACTTTCACAGGCAATTC 1980  
QY 1981 ATGCGCTGTAAGTATAGTATAGTAAATTAAGCCCACTAATTTGACTGAATTCGAAA 2040  
DB 1981 ATGCGCTGTAAGTATAGTATAGTAAATTAAGCCCACTAATTTGACTGAATTTGAAA 2040  
QY 2041 TTGATATGTTCTTACAGAGTGAAGATAAAGAAAAAGTCAACCAATGCGAGTCA 2100  
DB 2041 TTGATATGTTCTTACAGAGTGAAGATAAAGAAAAAGTCAACCAATGCGAGTCA 2100  
QY 2101 GGCACAGCAAGAACTAACAATCATGAGAGGTAAAGAACTGCAACTGAGCCAAAGAA 2160

DB 2101 GGCACAGCAAGAACTAACAATCATGAGAGGTAAAGAACTGCAACTGAGCCAAAGAA 2160  
QY 2161 GTAAACAGCCAAATGAAACAGACAGTAAGACACTGACAGTACTTTCCAGAGCTGA 2220  
DB 2161 GTAAACAGCCAAATGAAACAGACAGTAAGACACTGACAGTACTTTCCAGAGCTGA 2220  
QY 2221 AGTTAAACAAATGCACTGGTCTTTTACTAAGTGTTCCAATPACAGTAACCTTAAGAA 2280  
DB 2221 AGTTAAACAAATGCACTGGTCTTTTACTAAGTGTTCCAATPACAGTAACCTTAAGAA 2280  
QY 2281 TTGTCAATCTAGGCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTGT 2340  
DB 2281 TTGTCAATCTAGGCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTGT 2340  
QY 2341 CTATATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTGCAACTG 2400  
DB 2341 CTATATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTGCAACTG 2400  
QY 2401 AAGATCTGTAAGAGATGAGATATTTCAATGGTACCTGGTACTGATTAAGCACTAGG 2460  
DB 2401 AAGATCTGTAAGAGATGAGATATTTCAATGGTACCTGGTACTGATTAAGCACTAGG 2460  
QY 2461 AAGATATCTGTTACTGGAAGTTAGCACTAGGGAAGGCAAAAACAGAACTAAT 2520  
DB 2461 AAGATATCTGTTACTGGAAGTTAGCACTAGGGAAGGCAAAAACAGAACTAAT 2520  
QY 2521 GTGTGAGTCAAGTGTGACAGCATTTGAAAACCCCAAGGAGCACTAATGCTTCCAAAG 2580  
DB 2521 GTGTGAGTCAAGTGTGACAGCATTTGAAAACCCCAAGGAGCACTAATGCTTCCAAAG 2580  
QY 2581 ATATATGAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAGATTAACACAGTC 2640  
DB 2581 ATATATGAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAGATTAACACAGTC 2640  
QY 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGATGCTATTTGAGAAATCAT 2700  
DB 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGATGCTATTTGAGAAATCAT 2700  
QY 2701 TCAAGTTTCAAAAGCGGCACTCATTTGCTCTGTTTCAATCCAGGAAATGCAAGAGG 2760  
DB 2701 TCAAGTTTCAAAAGCGGCACTCATTTGCTCTGTTTCAATCCAGGAAATGCAAGAGG 2760  
QY 2761 AATGTGCAACATTTCTGCGCCACTCTGGTCTTAAAGAAACAAAGTCAAGTCACT 2820  
DB 2761 AATGTGCAACATTTCTGCGCCACTCTGGTCTTAAAGAAACAAAGTCAAGTCACT 2820  
QY 2821 TTGAATGTGAACAAAAGAGAAATCAAGGAAAGTGAATGATCAAGGCTGTAC 2880  
DB 2821 TTGAATGTGAACAAAAGAGAAATCAAGGAAAGTGAATGATCAAGGCTGTAC 2880  
QY 2881 AGACAGTTAATATACACGAGGCTTCTGTTGTCAGAAAGATTAACCCAGTTGATA 2940  
DB 2881 AGACAGTTAATATACACGAGGCTTCTGTTGTCAGAAAGATTAACCCAGTTGATA 2940  
QY 2941 ATGCGAAATGTAGATCAAGAGAGGCTCTAGGTTTTCTCATATCCTCAGTTCAGAGCA 3000  
DB 2941 ATGCGAAATGTAGATCAAGAGAGGCTCTAGGTTTTCTCATATCCTCAGTTCAGAGCA 3000  
QY 3001 ACAGAACTGACATCTACTTCAAAATTAACATGAGACTTTTACAAAACCCATATGATATAC 3060  
DB 3001 ACAGAACTGACATCTACTTCAAAATTAACATGAGACTTTTACAAAACCCATATGATATAC 3060  
QY 3061 CACGACCTTTTCCCATCAAGTCAATTTGTTAAACTAATATGAAGAAATCTGTAAGG 3120  
DB 3061 CACGACCTTTTCCCATCAAGTCAATTTGTTAAACTAATATGAAGAAATCTGTAAGG 3120  
QY 3121 AAAACCTTGAGAAACATTCATATGTCACCTGAAGAGAAAGGGAATGAGAACTTCAC 3180  
DB 3121 AAAACCTTGAGAAACATTCATATGTCACCTGAAGAGAAAGGGAATGAGAACTTCAC 3180  
QY 3181 GTACAGTGAACAAATTAAGCCGTAATTAACATTAAGAGAAATGTTTTAAAGAAAGCCAGCT 3240  
DB 3181 GTACAGTGAACAAATTAAGCCGTAATTAACATTAAGAGAAATGTTTTAAAGAAAGCCAGCT 3240



Db 3181 GTACAGTGAACACATTTAGCCGTAATTAACATTTAGAGAAAAATGTTTTTAAAGACCACT 3240  
OY 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTATGA 3300  
Db 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTATGA 3300  
OY 3301 TAGGTTCCAGTGTGAAAACATTTCAAGCAGACTAGTAGAAAACAGAGGCCAAATTTGA 3360  
Db 3301 TAGGTTCCAGTGTGAAAACATTTCAAGCAGACTAGTAGAAAACAGAGGCCAAATTTGA 3360  
OY 3361 ATGCTATGCTTAATATGAGGGTTTTGCAACCTGAGCTATTAACAAGTCTTCCTGGAA 3420  
Db 3361 ATGCTATGCTTAATATGAGGGTTTTGCAACCTGAGCTATTAACAAGTCTTCCTGGAA 3420  
OY 3421 GTATTTGTAAAGCATCTCTGAATAAAAAGCAAGATATGAAGAAGTGTTCAGACTGTA 3480  
Db 3421 GTATTTGTAAAGCATCTCTGAATAAAAAGCAAGATATGAAGAAGTGTTCAGACTGTA 3480  
OY 3481 ATACAGATTTCTCTCCATATCTGATTTGAGATAACTTAGAACAGCCTATGGGAAATGATC 3540  
Db 3481 ATACAGATTTCTCTCCATATCTGATTTGAGATAACTTAGAACAGCCTATGGGAAATGATC 3540  
OY 3541 ATGCACTCAGGTTTGTCTGAGAACCTGATGACCTGTATAGATGATGAATTAAG 3600  
Db 3541 ATGCACTCAGGTTTGTCTGAGAACCTGATGACCTGTATAGATGATGAATTAAG 3600  
OY 3601 AAGATCTAGTTTGTCTGAAATAGCATTAAGAAAAGTTCTGCTTTTGAAGAAAGC 3660  
Db 3601 AAGATCTAGTTTGTCTGAAATAGCATTAAGAAAAGTTCTGCTTTTGAAGAAAGC 3660  
OY 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTACCCATACATTTGGCTCAG 3720  
Db 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTACCCATACATTTGGCTCAG 3720  
OY 3721 GTTACGGAAGAGGGGCAAGAAATAGAGTCTCAGAGAGAACTTCTAGTAGAGATG 3780  
Db 3721 GTTACGGAAGAGGGGCAAGAAATAGAGTCTCAGAGAGAACTTCTAGTAGAGATG 3780  
OY 3781 AAGAGCTTCCCTGCTCCAAACATTTTATTTGTAAGTAACAATATACCTTCTAGT 3840  
Db 3781 AAGAGCTTCCCTGCTCCAAACATTTTATTTGTAAGTAACAATATACCTTCTAGT 3840  
OY 3841 CTACTAGAGCATAGCACGTTGCTACGAGAGTCTGCTTAGAACACAGAGAAATTTAT 3900  
Db 3841 CTACTAGAGCATAGCACGTTGCTACGAGAGTCTGCTTAGAACACAGAGAAATTTAT 3900  
OY 3901 TATCATTTGAAGATATACCTTAATAGCATGACGTAACAGTAATTTGGCAAGGCATCTC 3960  
Db 3901 TATCATTTGAAGATATACCTTAATAGCATGACGTAACAGTAATTTGGCAAGGCATCTC 3960  
OY 3961 AAGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTTGTCTTTCACAGTGA 4020  
Db 3961 AAGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTTGTCTTTCACAGTGA 4020  
OY 4021 GTGATTTGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
Db 4021 GTGATTTGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
OY 4081 CCAAAACAATGAGCATCAGTCTGAAGCCAGGAGTGTGCTGAGTGACAAAGAAATGG 4140  
Db 4081 CCAAAACAATGAGCATCAGTCTGAAGCCAGGAGTGTGCTGAGTGACAAAGAAATGG 4140  
OY 4141 TTTCAAGATGATGAAGAAAGAGAAAGGCTTTGGAAGAAATTAATCAAGAGACCAAGCA 4200  
Db 4141 TTTCAAGATGATGAAGAAAGAGAAAGGCTTTGGAAGAAATTAATCAAGAGACCAAGCA 4200  
OY 4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTAGAGTGAAGAAAGGCTCTCTGAAG 4260  
Db 4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTAGAGTGAAGAAAGGCTCTCTGAAG 4260  
OY 4261 ACTGCTCAGGGCTATCCTCTCAGAGTGAATTTTAAACCACTCAGCAGAGGAGTACCATGC 4320  
Db 4261 ACTGCTCAGGGCTATCCTCTCAGAGTGAATTTTAAACCACTCAGCAGAGGAGTACCATGC 4320  
OY 4321 AACATAACCTGATTAAGCTCCAGCAGAAATGGCTGACACTAGAGCTGTGTAGAACAGC 4380  
Db 4321 AACATAACCTGATTAAGCTCCAGCAGAAATGGCTGACACTAGAGCTGTGTAGAACAGC 4380  
OY 4381 ATGGAGCCAGCCTTCTTACAGCTACCCCTTCCATCATTAAGTACCTTCTGCTTGAAG 4440  
Db 4381 ATGGAGCCAGCCTTCTTACAGCTACCCCTTCCATCATTAAGTACCTTCTGCTTGAAG 4440  
OY 4441 ACCGCGCAATTCAGAACAAAGCAGATCCAGAAAACAGTATTAATCTCAGAAAGTA 4500  
Db 4441 ACCGCGCAATTCAGAACAAAGCAGATCCAGAAAACAGTATTAATCTCAGAAAGTA 4500  
OY 4501 GTGAATTAACCTTATTAAGCAGAAATCCAGAAAGCCTTCTCTGACAGAGTTTGAAGTGTG 4560  
Db 4501 GTGAATTAACCTTATTAAGCAGAAATCCAGAAAGCCTTCTCTGACAGAGTTTGAAGTGTG 4560  
OY 4561 CAGATAGTCTTACAGTAAATTAAGAAACAGAGTGAAGAGTATCCCTTTAAT 4620  
Db 4561 CAGATAGTCTTACAGTAAATTAAGAAACAGAGTGAAGAGTATCCCTTTAAT 4620  
OY 4621 GCCATCATTAATGATGATGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 4680  
Db 4621 GCCATCATTAATGATGATGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 4680  
OY 4681 ACTACCATCTCAAGAGAGAGCTTATTAAGTGTGATGATGATGATGATGATGATGATGAT 4740  
Db 4681 ACTACCATCTCAAGAGAGAGCTTATTAAGTGTGATGATGATGATGATGATGATGATGAT 4740  
OY 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTTCAGAGGCAAGATCTAGAGGCA 4800  
Db 4741 AGTCTGGGCCACAGATTTGACGAAACATCTTACTTTCAGAGGCAAGATCTAGAGGCA 4800  
OY 4801 CCCCTTAACCTGGAATGGAATGAGCTCTCTCTGATGATGATGATGATGATGATGATGAT 4860  
Db 4801 CCCCTTAACCTGGAATGGAATGAGCTCTCTCTGATGATGATGATGATGATGATGATGAT 4860  
OY 4861 AAGACAGAGCCCAAGAGAGCTGCTGTTGGCAACATACATCTTCAACCTTGCATTTGA 4920  
Db 4861 AAGACAGAGCCCAAGAGAGCTGCTGTTGGCAACATACATCTTCAACCTTGCATTTGA 4920  
OY 4921 AAGTCCCAATTAAGAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCCTACTACTG 4980  
Db 4921 AAGTCCCAATTAAGAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCCTACTACTG 4980  
OY 4981 ATACTGCTGGTATTAATGCAATGGAAGAAAGTGAAGAGGAGGAGGAGGAGGAGGAGG 5040  
Db 4981 ATACTGCTGGTATTAATGCAATGGAAGAAAGTGAAGAGGAGGAGGAGGAGGAGGAGG 5040  
OY 5041 CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100  
Db 5041 CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100  
OY 5101 AATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160  
Db 5101 AATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160  
OY 5161 CTGAGAGACTACTCATGTTGTTATGAAGACAGATGCTGAGTTTGTGTGAGAGGAGC 5220  
Db 5161 CTGAGAGACTACTCATGTTGTTATGAAGACAGATGCTGAGTTTGTGTGAGAGGAGC 5220  
OY 5221 TGAATATTTTTCAGAAATTTGGGGAGGAAATGGTATGCTATTTCTGGTGACCC 5280  
Db 5221 TGAATATTTTTCAGAAATTTGGGGAGGAAATGGTATGCTATTTCTGGTGACCC 5280  
OY 5281 AGCTATTAAGAAAGAAAGTGTGAATGAGCAGATTTTGAAGTCAAGAGAGATGAG 5340  
Db 5281 AGCTATTAAGAAAGAAAGTGTGAATGAGCAGATTTTGAAGTCAAGAGAGATGAG 5340  
OY 5341 TCAATGAAGAAACCAAGAGTCCAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5400  
Db 5341 TCAATGAAGAAACCAAGAGTCCAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5400



QY 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCACACATGCCACAGATCACTGG 5460  
 |||||||  
 DB 5401 TCAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCACACATGCCACAGATCACTGG 5460  
 QY 5461 AATGATGATGACAGCTGTGTGCTCTTCTGTGGTGAAGAGACTTTCATTCACCTTG 5520  
 |||||||  
 DB 5461 AATGATGATGACAGCTGTGTGCTCTTCTGTGGTGAAGAGACTTTCATTCACCTTG 5520  
 QY 5521 GCACAGGTGTCACCCCAATTTGTGTTGTGAGCCAGATGCTGACAGAGACAAATGGCT 5580  
 |||||||  
 DB 5521 GCACAGGTGTCACCCCAATTTGTGTTGTGAGCCAGATGCTGACAGAGACAAATGGCT 5580  
 QY 5581 TCCATGCAATTTGGCAGAGATGTGTGAGCACCCTGTGTGACCCGAGATGGGTGTGGACA 5640  
 |||||||  
 DB 5581 TCCATGCAATTTGGCAGAGATGTGTGAGCACCCTGTGTGACCCGAGATGGGTGTGGACA 5640  
 QY 5641 GTGTAGACATCTTACCACTGCCAGAGCTGGACACCTACCTGATACCCAGATCCCCACA 5700  
 |||||||  
 DB 5641 GTGTAGACATCTTACCACTGCCAGAGCTGGACACCTACCTGATACCCAGATCCCCACA 5700  
 QY 5701 GCCACTACTGA 5711  
 |||||||  
 DB 5701 GCCACTACTGA 5711

RESULT 13  
 AA287995 standard; DNA; 5712 BP.

AA287995;  
 06-JUN-2000 (first entry)  
 BRCA1 gene sequence (Genbank Accn No: U14680).

BRCA1: tumour; prostate cancer; cytostatic; antiproliferative;  
 gene therapy; ds.  
 Homo sapiens.

Location/Qualifiers  
 120..5711  
 /\*lag= a  
 /transl\_except= (pos: 1452..1454, aa:Asp)  
 /transl\_except= (pos: 1482..1484, aa:Asp)  
 /transl\_except= (pos: 1581..1583, aa:Ser)  
 /transl\_except= (pos: 1899..1901, aa:Glu)  
 /transl\_except= (pos: 1902..1904, aa:Leu)  
 /transl\_except= (pos: 1905..1907, aa:Glu)  
 /transl\_except= (pos: 1908..1910, aa:Leu)  
 /transl\_except= (pos: 1911..1913, aa:Asp)  
 /transl\_except= (pos: 1914..1916, aa:Leu)  
 /transl\_except= (pos: 1917..1919, aa:Met)  
 /transl\_except= (pos: 4476..4478, aa:Val)  
 /transl\_except= (pos: 4479..4481, aa:Leu)  
 /transl\_except= (pos: 4482..4484, aa:Glu)  
 /transl\_except= (pos: 4533..4535, aa:Xaa)  
 /transl\_except= (pos: 4689..4691, aa:Pro)  
 /note= "Xaa= unknown"

CA2218197-A1.  
 05-JUN-1999.  
 12-DEC-1997; 97CA-2218197.  
 05-DEC-1997; 97US-0986106.  
 (UNIV ) UNIV WASHINGTON.  
 (UYVA-) UNIV VANDERBILT.  
 Robinson-Benton CL, Thompson ME, Holt JT, Jensen RA, Steiner MS,  
 King M;

XX WPI: 2000-238071/21.  
 DR P-PSDB: AA177818.  
 XX  
 PT New method of treatment and suppression of prostate cancer comprises  
 PT using the BRCA family of genes to decrease the growth rate of the tumor  
 PT  
 PS Claim 11; Page 98-104; 166pp; English.  
 XX  
 CC The invention relates to a method for suppressing the growth of a  
 CC prostate tumor in a mammal that comprises introducing to the tumor a  
 CC vector comprising a nucleic acid sequence encoding a BRCA family gene  
 CC product operatively linked to a promoter, where production of the BRCA  
 CC family gene product results in a decrease in the growth rate of the  
 CC tumor. The methods are used to suppress the growth of and also to treat  
 CC prostate cancer in a mammal where the tumor is gene-linked hereditary  
 CC prostate cancer or sporadic prostate cancer. The present sequence  
 CC represents a BRCA1 gene sequence (Genbank Accn No: U14680) that can be  
 CC administered for suppressing tumor.  
 XX  
 SQ Sequence 5712 BP; 1956 A; 1099 C; 1274 G; 1383 T; 0 other;

Query Match 99.9%; Score 5707.8; DB 21; Length 5712;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCTGCGTGAAGACTTCTCTGACACCCGACAGAGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
 |||||||  
 DB 1 AGCTGCGTGAAGACTTCTCTGACACCCGACAGAGCTGTGGGTTTCTCAGATTAAGTGGCC 60  
 QY CCTGCGCTCAGAGAGAGCCCTTACCCCTGCTGCTGGGTAAGTTCATTTGGAACAGAAAGAA 120  
 |||||||  
 DB CCTGCGCTCAGAGAGAGCCCTTACCCCTGCTGCTGGGTAAGTTCATTTGGAACAGAAAGAA 120  
 QY 61 CTGCGCTCAGAGAGAGCCCTTACCCCTGCTGCTGGGTAAGTTCATTTGGAACAGAAAGAA 120  
 |||||||  
 DB 61 CTGCGCTCAGAGAGAGCCCTTACCCCTGCTGCTGGGTAAGTTCATTTGGAACAGAAAGAA 120  
 QY 121 TGGATTTATCTGCTCTTGGCGCTTGAAGAGTACAAATGTCATTATGCTATGCAAGAAA 180  
 |||||||  
 DB 121 TGGATTTATCTGCTCTTGGCGCTTGAAGAGTACAAATGTCATTATGCTATGCAAGAAA 180  
 QY 121 TGGATTTATCTGCTCTTGGCGCTTGAAGAGTACAAATGTCATTATGCTATGCAAGAAA 180  
 |||||||  
 DB 121 TGGATTTATCTGCTCTTGGCGCTTGAAGAGTACAAATGTCATTATGCTATGCAAGAAA 180  
 QY 181 TCTTAGAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCTCCAAAGTGTACC 240  
 |||||||  
 DB 181 TCTTAGAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCTCCAAAGTGTACC 240  
 QY 181 TCTTAGAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCTCCAAAGTGTACC 240  
 |||||||  
 DB 181 TCTTAGAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCTCCAAAGTGTACC 240  
 QY 241 ACATATTTTGAATTTTGAATGCTGAACCTTCAACCCAGAAAGAGGCTTCAAGT 300  
 |||||||  
 DB 241 ACATATTTTGAATTTTGAATGCTGAACCTTCAACCCAGAAAGAGGCTTCAAGT 300  
 QY 301 GTCTTTATGTAAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
 |||||||  
 DB 301 GTCTTTATGTAAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
 QY 301 GTCTTTATGTAAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
 |||||||  
 DB 301 GTCTTTATGTAAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
 QY 361 AACTTGTGAAGAGCTTGAATTAATCTGCTTTCAAGCTTGAACAGGTTTGGAGT 420  
 |||||||  
 DB 361 AACTTGTGAAGAGCTTGAATTAATCTGCTTTCAAGCTTGAACAGGTTTGGAGT 420  
 QY 421 ATGCAACAGCTATATTTTGAAGAAAGAAATTAATCTGCGAACAATCTAAAGATG 480  
 |||||||  
 DB 421 ATGCAACAGCTATATTTTGAAGAAAGAAATTAATCTGCGAACAATCTAAAGATG 480  
 QY 481 AAGTTTCTATCATTCAAGATATGGGCTACAGAAACGCTGCCAAAGACTTCTACAGAGT 540  
 |||||||  
 DB 481 AAGTTTCTATCATTCAAGATATGGGCTACAGAAACGCTGCCAAAGACTTCTACAGAGT 540  
 QY 541 AACCCGAAATCTCTCTTGCAGAGAAACGCTCACTGTCACAGTCTTAACCTTGAA 600  
 |||||||  
 DB 541 AACCCGAAATCTCTCTTGCAGAGAAACGCTCACTGTCACAGTCTTAACCTTGAA 600  
 QY 601 CTGTGAGACTCTGAGAGCAAGAGAGGATACCACTCAAAAGAGCTGTGTACTTG 660  
 |||||||  
 DB 601 CTGTGAGACTCTGAGAGCAAGAGAGGATACCACTCAAAAGAGCTGTGTACTTG 660  
 QY 661 AATGGATCTGATTTCTTGAAGATACCGTTATTAAGGCACTTATGCAAGTGTGGAG 720  
 |||||||  
 DB 661 AATGGATCTGATTTCTTGAAGATACCGTTATTAAGGCACTTATGCAAGTGTGGAG 720

Db 661 AATTGGATCTGATCTCTTGGAAGATACGCTTAATAAGCACTTATTGCACTGTGGAG 720  
Qy 721 ATCAGAATTTGTACAATCACCCTCAAGAACCAAGGATGAATCAGTTGGATCTG 780  
Db 721 ATCAGAATTTGTACAATCACCCTCAAGAACCAAGGATGAATCAGTTGGATCTG 780  
Qy 781 CAAAAAAGCTGCTTGTGAATTTTCTGAGACGATGTAACTAATCTGAACATCATCAC 840  
Db 781 CAAAAAAGCTGCTTGTGAATTTTCTGAGACGATGTAACTAATCTGAACATCATCAC 840  
Qy 841 CCAGTATATATGTTTGAACACACATGAGAGGCTGAGAGGATCCAGAAAGT 900  
Db 841 CCAGTATATATGTTTGAACACACATGAGAGGCTGAGAGGATCCAGAAAGT 900  
Qy 901 ATCAGGATGTTCTGTTTCAAACTGATGAGGACATGTGACACAATATCTATGCA 960  
Db 901 ATCAGGATGTTCTGTTTCAAACTGATGAGGACATGTGACACAATATCTATGCA 960  
Qy 961 GCTCATTTACAGCATGAGAACAGCATTTTATTCTCACTAAAGACAGATGATGAGAA 1020  
Db 961 GCTCATTTACAGCATGAGAACAGCATTTTATTCTCACTAAAGACAGATGATGAGAA 1020  
Qy 1021 AGGCTGAATTTCTGTATTAATAAGCAACAGCTGCTAGACAGGACCATTAACGAT 1080  
Db 1021 AGGCTGAATTTCTGTATTAATAAGCAACAGCTGCTAGACAGGACCATTAACGAT 1080  
Qy 1081 GGGCTGGAAGTAAGAAACATGTAATGATAGGCGGATCCACAGACAGAAAAAGGTAG 1140  
Db 1081 GGGCTGGAAGTAAGAAACATGTAATGATAGGCGGATCCACAGACAGAAAAAGGTAG 1140  
Qy 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGAATTAAGCAGAACTGCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGAATTAAGCAGAACTGCCATGCT 1200  
Qy 1201 CAGAGATCTGAGATGATGAGATGTTCTGATTAACATTAATAGAGATTCAGA 1260  
Db 1201 CAGAGATCTGAGATGATGAGATGTTCTGATTAACATTAATAGAGATTCAGA 1260  
Qy 1261 AAGTTAATGATGCTTTTCCAGAAAGTATGATGTTAGGTTCTGATGACTCATGATG 1320  
Db 1261 AAGTTAATGATGCTTTTCCAGAAAGTATGATGTTAGGTTCTGATGACTCATGATG 1320  
Qy 1321 GGGAGCTGATTAATGCCAAAGTATGATGATTTGGACGTTCTTAATGAGGTAGATG 1380  
Db 1321 GGGAGCTGATTAATGCCAAAGTATGATGATTTGGACGTTCTTAATGAGGTAGATG 1380  
Qy 1381 AATATCTGCTTCTGAGAAATAGACTTACTGCGACATGCTCATGAGGCTTTAA 1440  
Db 1381 AATATCTGCTTCTGAGAAATAGACTTACTGCGACATGCTCATGAGGCTTTAA 1440  
Qy 1441 TATGTAAGAGTGAAGAGTTCACCTCAATCAGTAGAGATTAATTGAAGACAAATAT 1500  
Db 1441 TATGTAAGAGTGAAGAGTTCACCTCAATCAGTAGAGATTAATTGAAGACAAATAT 1500  
Qy 1501 TTGGGAAAACTATCGAAGAGGCAAGCTCCCAACTTAAGCCATGTAACGAAATC 1560  
Db 1501 TTGGGAAAACTATCGAAGAGGCAAGCTCCCAACTTAAGCCATGTAACGAAATC 1560  
Qy 1561 TAATTTATAGAGATTTGTTACTGAGCAGATATATACAGAGGCGCCCTCACAATA 1620  
Db 1561 TAATTTATAGAGATTTGTTACTGAGCAGATATATACAGAGGCGCCCTCACAATA 1620  
Qy 1621 AATTAAGCGTAAAGAGACCTACATCAGGCTTCATCTGAGATTTTATCAAGAAAG 1680  
Db 1621 AATTAAGCGTAAAGAGACCTACATCAGGCTTCATCTGAGATTTTATCAAGAAAG 1680  
Qy 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGACTTAACCAAGGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGACTTAACCAAGGAGC 1740  
Qy 1741 AGAATGTCAGTGAATTAATTAATAGTGTATGATGATTAATAACCAAAAGGTGAT 1800  
Db 1741 AGAATGTCAGTGAATTAATTAATAGTGTATGATGATTAATAACCAAAAGGTGAT 1800  
Qy 1801 CTATTCAGATGAGAAAAATCTTAACCAATAGAAATCACTCGAAAAAGAAATCTGTTCA 1860  
Db 1801 CTATTCAGATGAGAAAAATCTTAACCAATAGAAATCACTCGAAAAAGAAATCTGTTCA 1860  
Qy 1861 AAACGAAGCTGAACTATTAAGCAGATTAAGCAATGTGAACTCGAATTAATTAATTC 1920  
Db 1861 AAACGAAGCTGAACTATTAAGCAGATTAAGCAATGTGAACTCGAATTAATTAATTC 1920  
Qy 1921 ACAATTCAAAAGCACTTAATAAGATAGGCTGAGGAAAGTCTTACACGATTAATTC 1980  
Db 1921 ACAATTCAAAAGCACTTAATAAGATAGGCTGAGGAAAGTCTTACACGATTAATTC 1980  
Qy 1981 ATGCGCTTGAACATGATAGTACAGTAAGATCTAAGCCACCTAATTTGATGAAATTC 2040  
Db 1981 ATGCGCTTGAACATGATAGTACAGTAAGATCTAAGCCACCTAATTTGATGAAATTC 2040  
Qy 2041 TTGATAGTTTCTTACAGTGAAGACATTAAGAAAAAAGTAAACCAATGCGAGTCA 2100  
Db 2041 TTGATAGTTTCTTACAGTGAAGACATTAAGAAAAAAGTAAACCAATGCGAGTCA 2100  
Qy 2101 GGCACAGCAGAACTTAACAATCATGGAAGGTAAAGACCTGCACTGAGCCAGAAAG 2160  
Db 2101 GGCACAGCAGAACTTAACAATCATGGAAGGTAAAGACCTGCACTGAGCCAGAAAG 2160  
Qy 2161 GTAAACAAGCCAAATGAAACAGACAGTAAGAAACATGACAGTATCTTCCAGAGCTGA 2220  
Db 2161 GTAAACAAGCCAAATGAAACAGACAGTAAGAAACATGACAGTATCTTCCAGAGCTGA 2220  
Qy 2221 AGTTAAACAATGACCTGCTCTTTTACTTAAGTGTTCATTAATCCAGTCACTTAAGAT 2280  
Db 2221 AGTTAAACAATGACCTGCTCTTTTACTTAAGTGTTCATTAATCCAGTCACTTAAGAT 2280  
Qy 2281 TTGTCAATCCTACCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGT 2340  
Db 2281 TTGTCAATCCTACCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGT 2340  
Qy 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAACTG 2400  
Db 2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAACTG 2400  
Qy 2401 AAAGATCTGTAAGAGATGACAGTATTTCAATGCTGTAACCTGTAATTAAGGACTCAG 2460  
Db 2401 AAAGATCTGTAAGAGATGACAGTATTTCAATGCTGTAACCTGTAATTAAGGACTCAG 2460  
Qy 2461 AAAGTATCTGTAAGAGATGACAGTATTTCAATGCTGTAACCTGTAATTAAGGACTCAG 2520  
Db 2461 AAAGTATCTGTAAGAGATGACAGTATTTCAATGCTGTAACCTGTAATTAAGGACTCAG 2520  
Qy 2521 GTGTAGTCAAGTGTGACAGATTTGAAAACCCCAAGGACTAATTAATGATGTTTCCAAG 2580  
Db 2521 GTGTAGTCAAGTGTGACAGATTTGAAAACCCCAAGGACTAATTAATGATGTTTCCAAG 2580  
Qy 2581 ATTAATGAATGACACAGAAAGCTTAAGTATTCATTTGGGACATTAAGTAAACCAAGT 2640  
Db 2581 ATTAATGAATGACACAGAAAGCTTAAGTATTCATTTGGGACATTAAGTAAACCAAGT 2640  
Qy 2641 GGGAAACAAGCATGAAATGGAAGAAAGTGAATGATGCTGATTAATTTGAGAAATACAT 2700  
Db 2641 GGGAAACAAGCATGAAATGGAAGAAAGTGAATGATGCTGATTAATTTGAGAAATACAT 2700  
Qy 2701 TCAAGTTTCAAGGCCAGCATTTGCTGTTTCAATCCAGAAATGACAGAAAGG 2760  
Db 2701 TCAAGTTTCAAGGCCAGCATTTGCTGTTTCAATCCAGAAATGACAGAAAGG 2760  
Qy 2761 AATGTCGAACATTTCTGCCCATCTGGGTCTTAAAGAACAAAGTCAAAAGTCACTT 2820  
Db 2761 AATGTCGAACATTTCTGCCCATCTGGGTCTTAAAGAACAAAGTCAAAAGTCACTT 2820  
Qy 2821 TTGAATGTGAACAAAGAAAGAAATCAAGGAAAGAAAGTGAATGATTAATCAAGCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAAAGAAATCAAGGAAAGAAAGTGAATGATTAATCAAGCTGTAC 2880

```

2881 AGACGTTAATATCATCTGACAGGCTTTCTGTGTGGTGTGACAGAAATAGCCAGTGTATA 2940
2881 AGACAGTTAATATCATCTGACAGGCTTTCTGTGTGGTGTGACAGAAATAGCCAGTGTATA 2940
2941 ATGCGCAATGTATGATCAAGAGAGGCTCTAGGTTTGTATCATCTCAGTTCAGAGCA 3000
2941 ATGCGCAATGTATGATCAAGAGAGGCTCTAGGTTTGTATCATCTCAGTTCAGAGCA 3000
3001 ACGAAATGACATCTTACCTCAAAATTAACATGAGCTTTTCAAAACCCATATGATATAC 3060
3001 ACGAAATGACATCTTACCTCAAAATTAACATGAGCTTTTCAAAACCCATATGATATAC 3060
3061 CACCACTTTTCCATCATAGTCTATTTGTTAAACTTAATGTAAGAAAATCTGCTAGAG 3120
3061 CACCACTTTTCCATCATAGTCTATTTGTTAAACTTAATGTAAGAAAATCTGCTAGAG 3120
3121 AAAACTTTGAGAACTTCAATGTCACCTGAAAGAAATGGGAAATGAGAACTATCCAA 3180
3121 AAAACTTTGAGAACTTCAATGTCACCTGAAAGAAATGGGAAATGAGAACTATCCAA 3180
3181 GTACATGAGCACAATTAGCCGTATTAACATTTAGAGAAATGTTTTAAAGACCAAGCT 3240
3181 GTACATGAGCACAATTAGCCGTATTAACATTTAGAGAAATGTTTTAAAGACCAAGCT 3240
3241 CAAGCAATATTATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3300
3241 CAAGCAATATTATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3300
3301 TAGGTTCCAGTATGAAAAACATTCAGACGAACTAGTAGAACAAGAGGCCAAATTTGA 3360
3301 TAGGTTCCAGTATGAAAAACATTCAGACGAACTAGTAGAACAAGAGGCCAAATTTGA 3360
3361 ATGCTATGCTAGATTGAGTTGGGTTTGCACCTGAGGCTATTAACAAAGCTTCTCGAA 3420
3361 ATGCTATGCTAGATTGAGTTGGGTTTGCACCTGAGGCTATTAACAAAGCTTCTCGAA 3420
3421 GTAATGTAAAGCACTCTGAAATATAAAAGCAAGATATGAAGAGTAGTTCAGACTGTA 3480
3421 GTAATGTAAAGCACTCTGAAATATAAAAGCAAGATATGAAGAGTAGTTCAGACTGTA 3480
3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAACCTTAGAACAGCCCTTAGGAGTAGTC 3540
3481 ATACAGATTTCTCTCCATATCTGATTTTCAGATTAACCTTAGAACAGCCCTTAGGAGTAGTC 3540
3541 ATGATCTCAGGTTTGTGTGAGACACCTGATGACCTGTTAGATGAGTGAATTAAGG 3600
3541 ATGATCTCAGGTTTGTGTGAGACACCTGATGACCTGTTAGATGAGTGAATTAAGG 3600
3601 AAGATACCTAGTTTGTGAAATGACATTAAGAAAGTTCTGCTGTTTTAGCAAAAAGCG 3660
3601 AAGATACCTAGTTTGTGAAATGACATTAAGAAAGTTCTGCTGTTTTAGCAAAAAGCG 3660
3661 TCCGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACATTTGGCTCAGG 3720
3661 TCCGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACATTTGGCTCAGG 3720
3721 GTTACGGAAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTTATCTAGTAGAGATG 3780
3721 GTTACGGAAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTTATCTAGTAGAGATG 3780
3781 AAGAGCTTCCCTGCTTCCAACTGTTATTTGTTAAAGTAAACAAATACCTTCTCAGT 3840
3781 AAGAGCTTCCCTGCTTCCAACTGTTATTTGTTAAAGTAAACAAATACCTTCTCAGT 3840
3841 CTACTAGCAGTAGCAGCCGTGCTACGAGTCTGCTAGAAACACAGAGAGAAATTTAT 3900
3841 CTACTAGCAGTAGCAGCCGTGCTACGAGTCTGCTAGAAACACAGAGAGAAATTTAT 3900
3901 TATCATGGAAGAAATAGCTTAATGATGACGTAACCGGTAATTTGGCAAAAGCACTTC 3960
3901 TATCATGGAAGAAATAGCTTAATGATGACGTAACCGGTAATTTGGCAAAAGCACTTC 3960
3961 AGGAACATCACTTAGTAGAGAAACAAATGTTCTGCTAGCTGTTTTCTTCACAGTGCA 4020
3961 AGGAACATCACTTAGTAGAGAAACAAATGTTCTGCTAGCTGTTTTCTTCACAGTGCA 4020

3961 AGGAACATCACTTAGTAGAGAAACAAATGTTCTGCTAGCTGTTTTCTTCACAGTGCA 4020
4021 GTGAATTTGGAAGACTTGAAGTCAAAATTAACAAACCCAGAGATCTTTTCTGATGTTCTT 4080
4021 GTGAATTTGGAAGACTTGAAGTCAAAATTAACAAACCCAGAGATCTTTTCTGATGTTCTT 4080
4081 CCAAAATGAGGATCAGTCTGAAAGCCAGGGAGTTGGTGTGAGTACAGAAAGATTTGG 4140
4081 CCAAAATGAGGATCAGTCTGAAAGCCAGGGAGTTGGTGTGAGTACAGAAAGATTTGG 4140
4141 TTTGAGATGATGAAGAAAGAGAGAGGCTTGGAAAGAAATATATCAAGAAAGCAAGCA 4200
4141 TTTGAGATGATGAAGAAAGAGAGAGGCTTGGAAAGAAATATATCAAGAAAGCAAGCA 4200
4201 TGGATTTCAAACTTGTGTGAGTACAGTCTGAGTGTGAGAGTGAAGAACGCTCTCTGAG 4260
4201 TGGATTTCAAACTTGTGTGAGTACAGTCTGAGTGTGAGAGTGAAGAACGCTCTCTGAG 4260
4261 ACTGCTCAGGCTATCTCTCTGAGAGTACATTTTAAACCACTCAGCAGAGGATACATGC 4320
4261 ACTGCTCAGGCTATCTCTCTGAGAGTACATTTTAAACCACTCAGCAGAGGATACATGC 4320
4321 AACATTAACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4380
4321 AACATTAACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4380
4381 ATGGAGGACCAAGCTTCTTAACAGCTACCTTCCATCATTAAGTACTTCTCCCTTGAGG 4440
4381 ATGGAGGACCAAGCTTCTTAACAGCTACCTTCCATCATTAAGTACTTCTCCCTTGAGG 4440
4441 ATGGAGGACCAAGCTTCTTAACAGCTACCTTCCATCATTAAGTACTTCTCCCTTGAGG 4440
4441 ATGGAGGACCAAGCTTCTTAACAGCTACCTTCCATCATTAAGTACTTCTCCCTTGAGG 4440
4441 ACCTGCGAAATCCAGAACCAAGCAATCAGAAAGAGAGTATTAACCTTCAACAGAAAGTA 4500
4441 ACCTGCGAAATCCAGAACCAAGCAATCAGAAAGAGAGTATTAACCTTCAACAGAAAGTA 4500
4501 GTGAATACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4560
4501 GTGAATACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4560
4561 GTGAATACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4560
4561 GTGAATACCTGATTAAGCTCCAGCAGAGAAATGAGTGAAGTGTGTTAGAACAGC 4560
4561 CAGATAGTTCTTACCAAGTAAATTAAGAACAGGAGTGAAGAGTCAATCCCTTCTAAT 4620
4561 CAGATAGTTCTTACCAAGTAAATTAAGAACAGGAGTGAAGAGTCAATCCCTTCTAAT 4620
4621 GCCCATCATTAAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4680
4621 GCCCATCATTAAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4680
4681 ACTACCATCTCAAGAGAGAGCTCAATTAAGTGTGTTGAATGTGAGAGCAACAGCTGGAAG 4740
4681 ACTACCATCTCAAGAGAGAGCTCAATTAAGTGTGTTGAATGTGAGAGCAACAGCTGGAAG 4740
4741 AGTCTGGGCCACAGATTTGACGGAACATCTTACTTCCAGAGAGATCTTAAGAGGAA 4800
4741 AGTCTGGGCCACAGATTTGACGGAACATCTTACTTCCAGAGAGATCTTAAGAGGAA 4800
4801 CCCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGAGACCTGGAATCTGATCTCTG 4860
4801 CCCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGAGACCTGGAATCTGATCTCTG 4860
4861 AAGACAGAGCCCAAGAGTCAAGCTGTTGGAACATTAACATCTTCAACCTCTCATTTGA 4920
4861 AAGACAGAGCCCAAGAGTCAAGCTGTTGGAACATTAACATCTTCAACCTCTCATTTGA 4920
4921 AAGTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCTGCTCATCTACTG 4980
4921 AAGTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCTGCTCATCTACTG 4980
4981 ATACTGCTGGGTATTAATGCAATGGAAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 5040
4981 ATACTGCTGGGTATTAATGCAATGGAAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 5040
5041 CTTTCACAGAAAGAGGTCAACAAAGATGTCATGAGTGTGCTGAGCTCAACCCAGAGAG 5100
5041 CTTTCACAGAAAGAGGTCAACAAAGATGTCATGAGTGTGCTGAGCTCAACCCAGAGAG 5100
```



421 ATGCAACGCTATATATTTGCAAAAAAGAAATTAAGTCTCTCGACATCTAAAGATG 480  
421 ATGCAACGCTATATATTTGCAAAAAAGAAATTAAGTCTCTCGACATCTAAAGATG 480  
481 AAGTTTCATGATCAAAAGTATGGCTACAGAAACCGGCAAAAGACTTCTACAGATG 540  
481 AAGTTTCATGATCAAAAGTATGGCTACAGAAACCGGCAAAAGACTTCTACAGATG 540  
541 AACCCGAAATCTTCCCTGAGAGAAACCACTCTAGTGTCAACTCTTAACCTTGGAA 600  
541 AACCCGAAATCTTCCCTGAGAGAAACCACTCTAGTGTCAACTCTTAACCTTGGAA 600  
601 CTGAGAGACTCTAGAGCAAAAGCAGGAGATCAACCTCAAAAGACGTCTGTACATG 660  
601 CTGAGAGACTCTAGAGCAAAAGCAGGAGATCAACCTCAAAAGACGTCTGTACATG 660  
661 AATTGGGATCTGATCTTCTGAGATACCTTAATTAAGGCAACTTATGCGTGGAG 720  
661 AATTGGGATCTGATCTTCTGAGATACCTTAATTAAGGCAACTTATGCGTGGAG 720  
721 ATCAAGATTTGTCAAAATCAACCCCTCAAGAGACCGAGATGAATCAGTTGGATTCTG 780  
721 ATCAAGATTTGTCAAAATCAACCCCTCAAGAGACCGAGATGAATCAGTTGGATTCTG 780  
781 CAAAAAGGCTGCTTGTGAATTTCTGAGAGGATGTAAACAATCTGAACATCATCAC 840  
781 CAAAAAGGCTGCTTGTGAATTTCTGAGAGGATGTAAACAATCTGAACATCATCAC 840  
841 CCAGTAATATGATTTTGACACCACTGGAAGCGTGCAGCTGAGAGCATCCAGAAAAGT 900  
841 CCAGTAATATGATTTTGACACCACTGGAAGCGTGCAGCTGAGAGCATCCAGAAAAGT 900  
901 ATCAGGAGATCTGTTTCAAACTGTCATGTGAGCCATGTGGCAAAATCTCATGCA 960  
901 ATCAGGAGATCTGTTTCAAACTGTCATGTGAGCCATGTGGCAAAATCTCATGCA 960  
961 GCTCATTACAGCATGAGAAAGCAGAGTTATTACTACTAAAGACAGAAATGAATGAGAAA 1020  
961 GCTCATTACAGCATGAGAAAGCAGAGTTATTACTACTAAAGACAGAAATGAATGAGAAA 1020  
1021 AGGCTGATTTCTGTAATTAAGCAAAAGCCTGGCTTAGCAAGAGCAACATAACAGAT 1080  
1021 AGGCTGATTTCTGTAATTAAGCAAAAGCCTGGCTTAGCAAGAGCAACATAACAGAT 1080  
1081 GGGCTGGAAGTAAAGAAATGTAATGATGATGAGCGGACTCCAGCAGCAAAAAAGGTAG 1140  
1081 GGGCTGGAAGTAAAGAAATGTAATGATGATGAGCGGACTCCAGCAGCAAAAAAGGTAG 1140  
1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATGAAGCAAACTGCCATGCT 1200  
1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATGAAGCAAACTGCCATGCT 1200  
1201 CAGGAATCTAGAGATTAAGATGTTCTTGGATTAACACTAATTAAGCAGCATTCGA 1260  
1201 CAGGAATCTAGAGATTAAGATGTTCTTGGATTAACACTAATTAAGCAGCATTCGA 1260  
1261 AAGTTAATGAGTGGTTTCCAGAGTGAATGATGATGATGATGATGATGATGATGATG 1320  
1261 AAGTTAATGAGTGGTTTCCAGAGTGAATGATGATGATGATGATGATGATGATGATG 1320  
1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATGAGAGCTTCTAATAGGTAGATG 1380  
1321 GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATGAGAGCTTCTAATAGGTAGATG 1380  
1381 AATTTCTGTTTCTGAGAGAAATAGACTTACTGGCAGTGAATCTCATAGGCTTTAA 1440  
1381 AATTTCTGTTTCTGAGAGAAATAGACTTACTGGCAGTGAATCTCATAGGCTTTAA 1440  
1441 TATGTAAAGAGTGAAGATTCATCTCAAAATAGTGAAGATTAATTTGAAGCAAAATAT 1500  
1441 TATGTAAAGAGTGAAGATTCATCTCAAAATAGTGAAGATTAATTTGAAGCAAAATAT 1500  
1501 TTGGGAAAAACCTATCGAAGAGCAAGCCTCCCAACTTAAGCATGTAACTGAAAAATC 1560  
1501 TTGGGAAAAACCTATCGAAGAGCAAGCCTCCCAACTTAAGCATGTAACTGAAAAATC 1560

1501 TTGGGAAAAACCTATCGAAGAGCAAGCCTCCCAACTTAAGCATGTAACTGAAAAATC 1560  
1561 TAATATGAGAGCATTTGTTACTGAGCAGATATATCAAGAGGTCCTCCCAAAATA 1620  
1561 TAATATGAGAGCATTTGTTACTGAGCAGATATATCAAGAGGTCCTCCCAAAATA 1620  
1621 AATTAAAGCGTAAAGAGACCTATACAGGCTTCACTGAGAGATTTTATCAAGAAAG 1680  
1621 AATTAAAGCGTAAAGAGACCTATACAGGCTTCACTGAGAGATTTTATCAAGAAAG 1680  
1621 AATTAAAGCGTAAAGAGACCTATACAGGCTTCACTGAGAGATTTTATCAAGAAAG 1680  
1621 AATTAAAGCGTAAAGAGACCTATACAGGCTTCACTGAGAGATTTTATCAAGAAAG 1680  
1661 CAGATTTGGCAGTTCAAAAGACTCTGTAATGATTAATCAGAGAACTAACCAACGGAGC 1740  
1661 CAGATTTGGCAGTTCAAAAGACTCTGTAATGATTAATCAGAGAACTAACCAACGGAGC 1740  
1661 CAGATTTGGCAGTTCAAAAGACTCTGTAATGATTAATCAGAGAACTAACCAACGGAGC 1740  
1661 CAGATTTGGCAGTTCAAAAGACTCTGTAATGATTAATCAGAGAACTAACCAACGGAGC 1740  
1741 AGAATGTCAGTGTATGATATTTACTATAGGTCATAGAGATTAATCAAAAGGTGAT 1800  
1741 AGAATGTCAGTGTATGATATTTACTATAGGTCATAGAGATTAATCAAAAGGTGAT 1800  
1741 AGAATGTCAGTGTATGATATTTACTATAGGTCATAGAGATTAATCAAAAGGTGAT 1800  
1741 AGAATGTCAGTGTATGATATTTACTATAGGTCATAGAGATTAATCAAAAGGTGAT 1800  
1801 CTATTCAGATGAGAAAAATCTTAACCCATAGATCACTGCAAAAAAGATCTGCTTCA 1860  
1801 CTATTCAGATGAGAAAAATCTTAACCCATAGATCACTGCAAAAAAGATCTGCTTCA 1860  
1861 AAACGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGAACTCGAATTAATATCC 1920  
1861 AAACGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGAACTCGAATTAATATCC 1920  
1921 ACAATTCAAAAGCAGCTTAAGAAAGATAGGAGAGAGAGAGAGAGAGAGAGAGATATTC 1980  
1921 ACAATTCAAAAGCAGCTTAAGAAAGATAGGAGAGAGAGAGAGAGAGAGAGAGATATTC 1980  
1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
1981 ATGCGCTTGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
2041 TTGATAGTTGTTTACAGAGTAAAGATTAAGAAAAAAGTAAACCAAAATGCCAGTCA 2100  
2041 TTGATAGTTGTTTACAGAGTAAAGATTAAGAAAAAAGTAAACCAAAATGCCAGTCA 2100  
2101 GGCACAGCAGAAACCTCAACTCATGGAAGTAAAGAACTCGCAACCAATATGCCAGTCA 2160  
2101 GGCACAGCAGAAACCTCAACTCATGGAAGTAAAGAACTCGCAACCAATATGCCAGTCA 2160  
2161 GTAACAGCCCAAAAGTAAAGCAGAGTAAAGCAATGACAGTGAATCTTCCAGAGCTGA 2220  
2161 GTAACAGCCCAAAAGTAAAGCAGAGTAAAGCAATGACAGTGAATCTTCCAGAGCTGA 2220  
2161 GTAACAGCCCAAAAGTAAAGCAGAGTAAAGCAATGACAGTGAATCTTCCAGAGCTGA 2220  
2161 GTAACAGCCCAAAAGTAAAGCAGAGTAAAGCAATGACAGTGAATCTTCCAGAGCTGA 2220  
2221 AGTTAACAAATGACACTGGTCTTCTTACTAGTGAATGATGATGATGATGATGATGATG 2280  
2221 AGTTAACAAATGACACTGGTCTTCTTACTAGTGAATGATGATGATGATGATGATGATG 2280  
2281 TTGTCAATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
2281 TTGTCAATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGATGATGATGATGATGATGATGATGATG 2400  
2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGATGATGATGATGATGATGATGATGATG 2400  
2401 AAAGATCTGAG 2460  
2401 AAAGATCTGAG 2460  
2461 AAAGATCTGAG 2520  
2461 AAAGATCTGAG 2520  
2521 GTGAGATCACTGAG 2580  
2521 GTGAGATCACTGAG 2580  
2581 ATTAATGAAGATGACAG 2640  
2581 ATTAATGAAGATGACAG 2640

Db 2581 ATATATGAAATGACACAGAAAGCTTTAATATCCATTGGGACATGAAGTTAACCACAGTC 2640  
QY 2641 GGGAAACACGATGGAATGGAAGAAAGTGAATGCTCAGTATTTGCGAATACAT 2700  
Db 2641 GGGAAACACGATGGAATGGAAGAAAGTGAATGCTCAGTATTTGCGAATACAT 2700  
QY 2701 TCAAGGTTTCAAGCGCCAGCTATTTGCTGTTTCAATCCAGGAATGCGAGAGAG 2760  
Db 2701 TCAAGGTTTCAAGCGCCAGCTATTTGCTGTTTCAATCCAGGAATGCGAGAGAG 2760  
QY 2761 AATGTGCAACATTTCTGCCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
Db 2761 AATGTGCAACATTTCTGCCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAAGTCACTT 2820  
QY 2821 TTGAATGTGAACAAGAAAGAAATCAAGAAAGATGAGTCAATATCAAGCTGTAC 2880  
Db 2821 TTGAATGTGAACAAGAAAGAAATCAAGAAAGATGAGTCAATATCAAGCTGTAC 2880  
QY 2881 AGACAGTTAATATCACTGACAGGCTTTCTGTTGGTCAAGAAAGATTAAGCCAGTTGATA 2940  
Db 2881 AGACAGTTAATATCACTGACAGGCTTTCTGTTGGTCAAGAAAGATTAAGCCAGTTGATA 2940  
QY 2941 ATGCCAAATGATGATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
Db 2941 ATGCCAAATGATGATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGAGCA 3000  
QY 3001 ACGAACACTGACATCTACTTCAAAATACATGACATTTTACAAACCCATATCTGTATAC 3060  
Db 3001 ACGAACACTGACATCTACTTCAAAATACATGACATTTTACAAACCCATATCTGTATAC 3060  
QY 3061 CACCACTTTTCCATCACTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG 3120  
Db 3061 CACCACTTTTCCATCACTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG 3120  
QY 3061 CACCACTTTTCCATCACTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG 3120  
Db 3061 CACCACTTTTCCATCACTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG 3120  
QY 3121 AAAACCTTTGGAAGCAATTCATGTCACCTGAAAGAAATGGGAAATGAGAACCTCCAA 3180  
Db 3121 AAAACCTTTGGAAGCAATTCATGTCACCTGAAAGAAATGGGAAATGAGAACCTCCAA 3180  
QY 3181 GTACAGTGAACCAATTAAGCCGTAAATACATTAAGAAAAATGTTTAAAGAACCCAGCT 3240  
Db 3181 GTACAGTGAACCAATTAAGCCGTAAATACATTAAGAAAAATGTTTAAAGAACCCAGCT 3240  
QY 3241 CAAGCAATATTAATGAAGTGAAGTTCAGTCAATTAAGAGTGGGCTCCAGATTAATGAA 3300  
Db 3241 CAAGCAATATTAATGAAGTGAAGTTCAGTCAATTAAGAGTGGGCTCCAGATTAATGAA 3300  
QY 3301 TAGCTTCAGTGAATGAAGCAATTCACAGCAAGTATGAGAAACAGAGGCGCAAAATGTA 3360  
Db 3301 TAGCTTCAGTGAATGAAGCAATTCACAGCAAGTATGAGAAACAGAGGCGCAAAATGTA 3360  
QY 3361 ATGCTATGCTTAGATTAGGGGTTTGCACCTGAGTCTATTAACAAGTCTTCTCGAA 3420  
Db 3361 ATGCTATGCTTAGATTAGGGGTTTGCACCTGAGTCTATTAACAAGTCTTCTCGAA 3420  
QY 3421 GTATATGTGAAGCATCTGGAATTAAGAAAGCAAGAAATGAAGAAATGAGTTCAGACTGTA 3480  
Db 3421 GTATATGTGAAGCATCTGGAATTAAGAAAGCAAGAAATGAAGAAATGAGTTCAGACTGTA 3480  
QY 3481 ATACAGTTTCTGCTCATCTGATTTCAATTAAGTGAAGCAAGCCATATGGAAGTGTGTC 3540  
Db 3481 ATACAGTTTCTGCTCATCTGATTTCAATTAAGTGAAGCAAGCCATATGGAAGTGTGTC 3540  
QY 3541 ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGGAATTAAGG 3600  
Db 3541 ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGGAATTAAGG 3600  
QY 3601 AAGATCTAGTTTGTCTGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAGCG 3660  
Db 3601 AAGATCTAGTTTGTCTGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAGCG 3660  
QY 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTACACCATACCATTTGGCTAGG 3720  
Db 3661 TCCAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTACACCATACCATTTGGCTAGG 3720

QY 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGAGTCTCAGAAAGAGACTTATCTAGTGAAGATG 3780  
Db 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGAGTCTCAGAAAGAGACTTATCTAGTGAAGATG 3780  
QY 3781 AAGAGCTTCCCTGCTTCAACACTTGTATTTTGTAAAGTAAACAATATACCTTCTCACT 3840  
Db 3781 AAGAGCTTCCCTGCTTCAACACTTGTATTTTGTAAAGTAAACAATATACCTTCTCACT 3840  
QY 3841 CTACTAGGCAATAGCAGCGTGTCTACCGAGTGTCTGTCTAAGAAACACAGAGGAAATTAAT 3900  
Db 3841 CTACTAGGCAATAGCAGCGTGTCTACCGAGTGTCTGTCTAAGAAACACAGAGGAAATTAAT 3900  
QY 3901 TATCATTAAGAAATAGCTTAAATGACGACAGTAAACAGGTAATATGGAAGGCAATCTAC 3960  
Db 3901 TATCATTAAGAAATAGCTTAAATGACGACAGTAAACAGGTAATATGGAAGGCAATCTAC 3960  
QY 3961 AGGAACATCACTTAGTGAAGAAACAAATGTTCTGCTAGCTTGTCTTCTACAGTGTCA 4020  
Db 3961 AGGAACATCACTTAGTGAAGAAACAAATGTTCTGCTAGCTTGTCTTCTACAGTGTCA 4020  
QY 4021 GTGAATTTGAAGAGCTTACTGCAATTAACAACCCAGAGATCTTCTGATTTGGTCTT 4080  
Db 4021 GTGAATTTGAAGAGCTTACTGCAATTAACAACCCAGAGATCTTCTGATTTGGTCTT 4080  
QY 4081 CCAAAACAAATGAGGATCACTGTAAGAACCCAGAGAGTGGTCTGAGTGAACAAGAAATGG 4140  
Db 4081 CCAAAACAAATGAGGATCACTGTAAGAACCCAGAGAGTGGTCTGAGTGAACAAGAAATGG 4140  
QY 4141 TTTGATGATGATGAAGAAAGAGAAACGGGCTTGAAGAAATTAATCAAGAAAGCAAGCA 4200  
Db 4141 TTTGATGATGATGAAGAAAGAGAAACGGGCTTGAAGAAATTAATCAAGAAAGCAAGCA 4200  
QY 4201 TGGATTTAACTTAAGTGAAGAGCAAGCTGGGTTGTGAAGTGAACCAAGCTCTGGAAG 4260  
Db 4201 TGGATTTAACTTAAGTGAAGAGCAAGCTGGGTTGTGAAGTGAACCAAGCTCTGGAAG 4260  
QY 4261 ACTGCTCAGGCTATCTCTCAGAGTGACATTTTAAACCACTGACAGAGAGATACCATTC 4320  
Db 4261 ACTGCTCAGGCTATCTCTCAGAGTGACATTTTAAACCACTGACAGAGAGATACCATTC 4320  
QY 4321 AACATTAACCTGATTAAGAGCTCAGCAGAGAAATGGCTGAAGTGAAGCTGTGTTGAACAGC 4380  
Db 4321 AACATTAACCTGATTAAGAGCTCAGCAGAGAAATGGCTGAAGTGAAGCTGTGTTGAACAGC 4380  
QY 4381 ATGGAGCCAGGCTTCTTAACAGCTACCCCTCCATTAAGTGAAGTGTGCTTGGCTTGAAG 4440  
Db 4381 ATGGAGCCAGGCTTCTTAACAGCTACCCCTCCATTAAGTGAAGTGTGCTTGGCTTGAAG 4440  
QY 4441 ACCTGCAAAATCCAGAAACAAAGCAATCAAGAAACAGTATTAATCTTCAAGAAAGTA 4500  
Db 4441 ACCTGCAAAATCCAGAAACAAAGCAATCAAGAAACAGTATTAATCTTCAAGAAAGTA 4500  
QY 4501 GTGAATTAACCTTAAGCAGATTCAGAAAGGCTTCTCTGCAAGAGTTGAGGTGTCTG 4560  
Db 4501 GTGAATTAACCTTAAGCAGATTCAGAAAGGCTTCTCTGCAAGAGTTGAGGTGTCTG 4560  
QY 4561 CAGATGTTCTTACAGTAAATAAGAAACAGGAGTGAAGAGTCAATCCCTTGAAT 4620  
Db 4561 CAGATGTTCTTACAGTAAATAAGAAACAGGAGTGAAGAGTCAATCCCTTGAAT 4620  
QY 4621 GCCCATCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
Db 4621 GCCCATCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4680  
QY 4681 ACTTACCATCTCAAGAGAGCTCATTAAGGTTTGTATGATGATGATGATGATGATGATGATG 4740  
Db 4681 ACTTACCATCTCAAGAGAGCTCATTAAGGTTTGTATGATGATGATGATGATGATGATGATG 4740  
QY 4741 AGTCTGGGCAACAGATTTGAGGAAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA 4800  
Db 4741 AGTCTGGGCAACAGATTTGAGGAAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA 4800

OY	4801	CCCCCTACCTGGAAATCGAGATCAAGCCCTCTCTCGATGACCCGGAATCTGATCCCTCTG	4860
Db	4801	CCCCCTACCTGGAAATCGAGATCAAGCCCTCTCTCGATGACCCGGAATCTGATCCCTCTG	4860
OY	4861	AAGCAGAGAGCCCCAGAGCTCAAGCTCGTTGGCAACATACATCTTCAACCTCTCATTTGA	4920
Db	4861	AAGCAGAGAGCCCCAGAGCTCAAGCTCGTTGGCAACATACATCTTCAACCTCTCATTTGA	4920
OY	4921	AAGTTCGCCCAATTGAAGTTGCAACAATCTGGCCAGAGTCCAGCTGTGCTCATACTCTG	4980
Db	4921	AAGTTCGCCCAATTGAAGTTGCAACAATCTGGCCAGAGTCCAGCTGTGCTCATACTCTG	4980
OY	4981	ATACTGCTGGGTATTAATGCATGGAAGAAGTGTGACGAGGAGAAAGCCAGATTTGACG	5040
Db	4981	ATACTGCTGGGTATTAATGCATGGAAGAAGTGTGACGAGGAGAAAGCCAGATTTGACG	5040
OY	5041	CTTCACAGAAAGGGTCAACAAAAGAAATGTCATGTGTGTCTGAGCCCTGACCCAGAG	5100
Db	5041	CTTCACAGAGAGGGTCAACAAAAGAAATGTCATGTGTGTCTGAGCCCTGACCCAGAG	5100
OY	5101	AATTTATGCTGTGTAGCAAGTTTGCAGAAAACACCAATCACTTAACTAATCTAAATTA	5160
Db	5101	AATTTATGCTGTGTAGCAAGTTTGCAGAAAACACCAATCACTTAACTAATCTAAATTA	5160
OY	5161	CTGAGAGAGCTACTCATGTTGTTATGAAGACAGATGCTGAATTTGTGTGTGACGAGAC	5220
Db	5161	CTGAGAGAGCTACTCATGTTGTTATGAAGACAGATGCTGAATTTGTGTGTGACGAGAC	5220
OY	5221	TGAAATATTTTCTAGGAATTCGGGAGAGAAATGGTATAGTATTTCTGGGCTGAGCC	5280
Db	5221	TGAAATATTTTCTAGGAATTCGGGAGAGAAATGGTATAGTATTTCTGGGCTGAGCC	5280
OY	5281	AGCTTATTTAAAGAAAGAAAAATGCTGATGATGCAATGATTTTGAAGTCAGAGAGATGTG	5340
Db	5281	AGCTTATTTAAAGAAAGAAAAATGCTGATGATGCAATGATTTTGAAGTCAGAGAGATGTG	5340
OY	5341	TCAATGGAAGAACCCACCAGGTGCCAAAGCGAGCAGAGAGAAATCCAGAGCAGAAAGATCT	5400
Db	5341	TCAATGGAAGAAACCCACCAGGTGCCAAAGCGAGCAGAGAGAAATCCAGAGCAGAAAGATCT	5400
OY	5401	TCAGGGGGCTTAGAATCTGTTGCTATGGCCCTTCACCACATGCCCAACATCAACTCG	5460
Db	5401	TCAGGGGGCTTAGAATCTGTTGCTATGGCCCTTCACCACATGCCCAACATCAACTCG	5460
OY	5461	AATGATGTGTACAGCTGTGTGCTCTCTGTGGTGAAGAGCTTTCATCATTTCAACCCCTG	5520
Db	5461	AATGATGTGTACAGCTGTGTGCTCTCTGTGGTGAAGAGCTTTCATCATTTCAACCCCTG	5520
OY	5521	GCACAGGTGTCCACCAATTGTGTTGTGTCAGCCAGATGCTGTGACAGAGAACAAATGGCT	5580
Db	5521	GCACAGGTGTCCACCAATTGTGTTGTGTCAGCCAGATGCTGTGACAGAGAACAAATGGCT	5580
OY	5581	TTCATGCAATTTGGGACAGATGTGTAGGACACTGTGTGTGATACCCGAGAGTGGGTGTTGACA	5640
Db	5581	TTCATGCAATTTGGGACAGATGTGTAGGACACTGTGTGTGATACCCGAGAGTGGGTGTTGACA	5640
OY	5641	GTTGAGCACTTACCAAGTGCAGAGAGTGGACACTTACCTGATATCCCAAGATCCCCACACA	5700
Db	5641	GTTGAGCACTTACCAAGTGCAGAGAGTGGACACTTACCTGATATCCCAAGATCCCCACACA	5700
OY	5701	GCCACTACTGA 5711	
Db	5701	GCCACTACTGA 5711	
RESULT 15			
AAAT17438 standard; cDNA: 5914 BP.			
XX	AAAT17438;		
XX	AAAT17438;		
DT	30-SEP-1996 (first entry)		

DE	BRCA1 coding sequence.
XX	Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW	antibody production; germline alteration; probe; lesion neoplasia; human;
KM	gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS 120..5711
FT	/tag= a
FT	/product= BRCA1
XX	
PN	W09605306-A2.
XX	
PD	22-FEB-1996.
XX	
PF	11-AUG-1995; 95WO-US10202.
XX	
PR	07-JUN-1995; 95US-0483553.
PR	12-AUG-1994; 94US-0289221.
PR	02-SEP-1994; 94US-0300266.
PR	16-SEP-1994; 94US-0308104.
PR	29-NOV-1994; 94US-0348824.
PR	24-MAR-1995; 95US-0409305.
PR	07-JUN-1995; 95US-0480784.
XX	
PA	(MYRI-) MYRIAD GENETICS INC.
PA	(CANC-) CANCER INST.
PA	(RECH-) CENT RECH DU CHUL.
PI	Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
DR	WPI: 1996-139702/14.
DR	P-PSDB: AAR81481.
XX	
PT	New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT	gene - for diagnosis and therapy of human breast and ovarian cancer
PT	and for diagnosing pre-disposition to these cancers
XX	
PS	Claim 1; Page 119-128; 218pp; English.
CC	This sequence represents the cDNA of the human breast and ovarian cancer
CC	predisposing gene (BRCA1). Proteins encoded by mutations of this
CC	sequence (see AAT17439-T17453 and AAT17455-T17529) can be used as
CC	immunogens for antibody production. An additional BRCA1 mutant, with a T
CC	to G mutation and 55 nucleotide insert (encoding a protein which
CC	terminates at residue 75) is also claimed, however the exact location of
CC	the mutations is not specified so the sequence could not be created. The
CC	mutant BRCA1 genes have at least 1 mutation or polymorphism in comparison
CC	to this sequence. By detecting a germline alteration in this gene, a
CC	predisposition for breast and ovarian cancer can be diagnosed. In one
CC	method, BRCA1 mRNA isolated from a tissue sample from a subject has a
CC	probe, corresponding to a fragment of this sequence (or an
CC	allele-specific probe for a mutation of this sequence), added to it.
CC	The conditions allow for hybridisation of the probe to the mRNA, and any
CC	hybridisation which occurs is detected. Alternatively the BRCA1 gene in
CC	the tissue sample is isolated, and a shift in electrophoretic mobility
CC	of single stranded DNA from the sample on a non-denaturing
CC	polyacrylamide gel indicates a mutation. These methods of detection can
CC	also diagnose a lesion neoplasia associated with the BRCA1 locus. The
CC	methods may be used in gene therapy, protein replacement therapy and
CC	protein mimetics, and may be used to screen for drugs in cancer therapy.
XX	
SO	Sequence 5914 BP; 2006 A; 1156 C; 1316 G; 1436 T; 0 other;
XX	
Query Match	99.9%; Score 5707.8; DB 17; Length 5914;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 5709;	Conservative 0; Mismatches 2; Indels 0; Gaps 0;
db	1 AGCTGCGTGAAGACTCTTCGAGCCCCGACACAGGCTGTGGGGTTTCTCAGATACTGGGCC 60
	1 AGCTGCGTGAAGACTCTTCGAGCCCCGACACAGGCTGTGGGGTTTCTCAGATACTGGGCC 60



QY	63	CGTGGCTCAGGAGGCGCTTACCCCTCTGCTCTGGGTAAGATTCATTGGACAGCAAGAA	120
Db	61	CGTGGCTCAGGAGGCGCTTACCCCTCTGCTCTGGGTAAGATTCATTGGACAGCAAGAA	120
QY	121	TGGATTTATCTCTCTGCTGCTTGGAAAGTACAAATGTCAATTAATGTTATGACAGAAA	180
Db	121	TGGATTTATCTCTCTGCTGCTTGGAAAGTACAAATGTCAATTAATGTTATGACAGAAA	180
QY	181	TCTTAGAGTGTCCCATCTGTCTGGAGTTGATATCAAGAACCTTCTTCACAAAGTGTACC	240
Db	181	TCTTAGAGTGTCCCATCTGTCTGGAGTTGATATCAAGAACCTTCTTCACAAAGTGTACC	240
QY	241	ACATATTTTGGCAAAATTTTGCATGTGCAAACTTCTCAACGAGAAGAGGCGCTTACAGT	300
Db	241	ACATATTTTGGCAAAATTTTGCATGTGCAAACTTCTCAACGAGAAGAGGCGCTTACAGT	300
QY	301	GTCCTTTATGTAGAATGATATTAACCAAAAGGAGCCCTACAGAAAGTACGAGATTTAGTC	360
Db	301	GTCCTTTATGTAGAATGATATTAACCAAAAGGAGCCCTACAGAAAGTACGAGATTTAGTC	360
QY	361	AACCTGTGAACAGGATATTGAAATCATTTGTGCTTTACGTTGACGTAACAGGTTGGAGT	420
Db	361	AACCTGTGTGAACAGCTATTGAAATCATTTGTGCTTTACGTTGACGTAACAGGTTGGAGT	420
QY	421	ATGCAACAGCTATTAATTTTGGCAAAAAGGAAATTAACCTCTCTCAACATCTTAAAGATG	480
Db	421	ATGCAACAGCTATTAATTTTGGCAAAAAGGAAATTAACCTCTCTCAACATCTTAAAGATG	480
QY	481	AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCACAAAGACTTCTACAGAGT	540
Db	481	AAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCACAAAGACTTCTACAGAGT	540
QY	541	AACCCGAATATCTTCTCTTGGCAGAGAACCAAGTACAGTGTCCACATCTCTAATCTTGGAA	600
Db	541	AACCCGAATATCTTCTCTTGGCAGAGAACCAAGTACAGTGTCCACATCTCTAATCTTGGAA	600
QY	601	CTGTAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCAATTTG	660
Db	601	CTGTAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCAATTTG	660
QY	661	AATTGGAGTCTGATCTTCTGAGAGTACCCTTAATTAAGGCAACTTAATGTGACGTGGGAG	720
Db	661	AATTGGAGTCTGATCTTCTGAGAGTACCCTTAATTAAGGCAACTTAATGTGACGTGGGAG	720
QY	721	ATCAAGAATTTGTTACAAATCAACCCCTCAAGGAACCGAGGATGAATTCAGATTGTG	780
Db	721	ATCAAGAATTTGTTACAAATCAACCCCTCAAGGAACCGAGGATGAATTCAGATTGTG	780
QY	781	CAAAAAGGCTGTGTAATTTTGTAGAGGAGTAAACAATCTGAACATCATCAAC	840
Db	781	CAAAAAGGCTGTGTAATTTTGTAGAGGAGTAAACAATCTGAACATCATCAAC	840
QY	841	CCAGTAATTAATGATTTGAACACCACTGAGGAAGCGTGAAGGATCCAGAAAGT	900
Db	841	CCAGTAATTAATGATTTGAACACCACTGAGGAAGCGTGAAGGATCCAGAAAGT	900
QY	901	ATCAGGGTAGTTCCTTTCAAACTTGCACTGTGGAGCATGTGGCACAATCTCATGCGCA	960
Db	901	ATCAGGGTAGTTCCTTTCAAACTTGCACTGTGGAGCATGTGGCACAATCTCATGCGCA	960
QY	961	GCTCATTTACACATGAGAACAGCAAGTTTATTACTCACTTAAGACAGATGATGAGAA	1020
Db	961	GCTCATTTACACATGAGAACAGCAAGTTTATTACTCACTTAAGACAGATGATGAGAA	1020
QY	1021	AGGCTGAATTTCTGTAATTAAGCAAAACAGCCCTGCTTGAACAAGGACCAATTAACAGAT	1080
Db	1021	AGGCTGAATTTCTGTAATTAAGCAAAACAGCCCTGCTTGAACAAGGACCAATTAACAGAT	1080
QY	1081	GGGCTGGAAGTAAAGAAACATGTATATATAGCGGACCTCCACAGACAGAAAAAAGTGTG	1140
Db	1081	GGGCTGGAAGTAAAGAAACATGTATATATAGCGGACCTCCACAGACAGAAAAAAGTGTG	1140

OY	1141	ATCTGAATCGTCATCCCTCGTGTGAGAGAAAGAAATGGAAATAGACAAACCTGCATCT	1200
Db	1141	ATCTGAATCGTCATCCCTCGTGTGAGAGAAAGAAATGGAAATAGACAAACCTGCATCT	1200
OY	1201	CAGAGAAATCCCTGAGATAGCTGAAGATGTTCTTGATTAACATCTAAATAGCAGCATTCAGA	1260
Db	1201	CAGAGAAATCCCTGAGATAGCTGAAGATGTTCTTGATTAACATCTAAATAGCAGCATTCAGA	1260
OY	1261	AAGTTAATGAGTGGTTTTCCAGAACTGATGAACCTGTAAAGTTCGTATGACTCCACATGATG	1320
Db	1261	AAGTTAATGAGTGGTTTTCCAGAACTGATGAACCTGTAAAGTTCGTATGACTCCACATGATG	1320
OY	1321	GGGACCTCGAATCAAAATGCCAAGTAGCTGATATTTGAGCTCTCTAAATGAGTAGATG	1380
Db	1321	GGGACCTCGAATCAAAATGCCAAGTAGCTGATATTTGAGCTCTCTAAATGAGTAGATG	1380
OY	1381	AATATTCCTGTTCTTCAGAGAAATAGACTTACTGGCCAGTAGATCTCATGAGCTTTTAA	1440
Db	1381	AATATTCCTGTTCTTCAGAGAAATAGACTTACTGGCCAGTAGATCTCATGAGCTTTTAA	1440
OY	1441	TATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATATATTTGAAGCAAAATAT	1500
Db	1441	TATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATATATTTGAAGCAAAATAT	1500
OY	1501	TTGGGAAAAACCTATCGGAAAGAGGAGAACCTCCCACTTAAAGCATGTAACTGAAAAATC	1560
Db	1501	TTGGGAAAAACCTATCGGAAAGAGGAGAACCTCCCACTTAAAGCATGTAACTGAAAAATC	1560
OY	1561	TAAATATAGAGACATTTTGTACTGAGCCACAGATATATCAAGAGAGTCCCTCACAAATA	1620
Db	1561	TAAATATAGAGACATTTTGTACTGAGCCACAGATATATCAAGAGAGTCCCTCACAAATA	1620
OY	1621	AATTTAAAGCGTAAAGAGAGACCTACATCAGGCCCTTATCTGAGAGTTTATCAAGAAG	1680
Db	1621	AATTTAAAGCGTAAAGAGAGACCTACATCAGGCCCTTATCTGAGAGTTTATCAAGAAG	1680
OY	1681	CAGATTTTGGCAGTTTAAAAAGACTCCTGTAATATGATTAATCAAGGAACTAAACAAAGCGAGC	1740
Db	1681	CAGATTTTGGCAGTTTAAAAAGACTCCTGTAATATGATTAATCAAGGAACTAAACAAAGCGAGC	1740
OY	1741	AGAATGGTCAAGTGTATGATATTTACTAATAGTGGCATGAGAAATTAACAAAGCTGAT	1800
Db	1741	AGAATGGTCAAGTGTATGATATTTACTAATAGTGGCATGAGAAATTAACAAAGCTGAT	1800
OY	1801	CTATTCAGAAATGAGAAAAATCCTTAACCCAAATAGATCACTCGAAAAAAGAAATCTGCTTCA	1860
Db	1801	CTATTCAGAAATGAGAAAAATCCTTAACCCAAATAGATCACTCGAAAAAAGAAATCTGCTTCA	1860
OY	1861	AAAGCAAAAGCTGACCTTAAAGCAGCACTATATAGCAATATGAGACTGCAATTAATATCC	1920
Db	1861	AAAGCAAAAGCTGACCTTAAAGCAGCACTATATAGCAATATGAGACTGCAATTAATATCC	1920
OY	1921	ACAATTTCAAAAGCACCTTAAAAAGAAATAGGCTGAGAGAGAGCTTCTACCGAGCATATTC	1980
Db	1921	ACAATTTCAAAAGCACCTTAAAAAGAAATAGGCTGAGAGAGAGCTTCTACCGAGCATATTC	1980
OY	1981	ATGGCCTTGAACCTAGTACTGATGTAAGAAATCTAAGGCCACCTAATTTGTACTAAATTCGAAA	2040
Db	1981	ATGGCCTTGAACCTAGTACTGATGTAAGAAATCTAAGGCCACCTAATTTGTACTAAATTCGAAA	2040
OY	2041	TTGATAGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACACCAATATGCCAGTCA	2100
Db	2041	TTGATAGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACACCAATATGCCAGTCA	2100
OY	2101	GGCAGCAGCAAACTTACACTCAATGGAAGGTAAAGAACTGCAACTGAGCCCAAGAAAGA	2160
Db	2101	GGCAGCAGCAAACTTACACTCAATGGAAGGTAAAGAACTGCAACTGAGCCCAAGAAAGA	2160
OY	2161	GTAACAAACCAATATCAAGACAACTAAAGACATGAAGTATACCTTTCCAGAGCTGA	2220
Db	2161	GTAACAAACCAATATCAAGACAACTAAAGACATGAAGTATACCTTTCCAGAGCTGA	2220
OY	2221	AGTTAACAATGACACCTGGTCTTTTACTAAGTGTCAAAATACCAAGTGAATTTAAAGAT	2280



Db	2221	AGTTAACAATAATGCACACCTGGTTCTTTACTAAGTGTCAATATACAGTAGAATTAAAGAT	2280
OY	2281	TTGTCAATCCTAGCCTTCCAAAGAGAGAAAAAGAGAAAGAACTAGAAACAGTTAAAGTGT	2340
Db	2281	TTGTCAATCCTAGCCTTCCAAAGAGAGAAAAAGAGAAAGAAAGAACTAGAAACAGTTAAAGTGT	2340
OY	2341	CTAATAATGCTGAAGAGACCCCAAGATGTCATGTTAAATGAGAAAGGGTTTTCGAAACTG	2400
Db	2341	CTAATAATGCTGAAGAGACCCCAAGATGTCATGTTAAATGAGAAAGGGTTTTCGAAACTG	2400
OY	2401	AAAGATCTGTAGAGAGTAGAGATTAATTCATTTGACTGCTGACTGATTTATGGCACTCAG	2460
Db	2401	AAAGATCTGTAGAGAGTAGAGATTTTCATTTGACTGCTGACTGATTTATGGCACTCAG	2460
OY	2461	AAATATCTCTGTTACTGGAAGTTAGCACTGTAGGGAAGGCAAAAACGAACCAAAATTAAT	2520
Db	2461	AAATATCTCTGTTACTGGAAGTTAGCACTGTAGGGAAGGCAAAAACGAACCAAAATTAAT	2520
OY	2521	GTGTGACTGCAGTGTGCAGCAATTTTGAAGAACCCCAAGGACATAATTCATGTTGTTCCAAAG	2580
Db	2521	GTGTGACTGCAGTGTGCAGCAATTTGGAAGAACCCCAAGGACATAATTCATGTTGTTCCAAAG	2580
OY	2581	ATAATAGAAATGACACAGAAAGGCTTTAAGTATTCATTTGGACATGAAGTTAACCACAGTC	2640
Db	2581	ATAATAGAAATGACACAGAAAGGCTTTAAGTATTCATTTGGACATGAAGTTAACCACAGTC	2640
OY	2641	GGGAACACAGCATGTGAANAATGGAACAAAGTGAACTGATGCTGATTTTCACAAATTCAT	2700
Db	2641	GGGAACACAGCATGTGAANAATGGAACAAAGTGAACTGATGCTGATTTTCACAAATTCAT	2700
OY	2701	TCAAGGTTTTCAAGGCGCAGTCATTTTGCTGTGTTTTCAATTCACGAAATGTCAGAAAGAGG	2760
Db	2701	TCAAGGTTTTCAAGGCGCAGTCATTTTGCTGTGTTTTCAATTCACGAAATGTCAGAAAGAGG	2760
OY	2761	AATGTGCACATCTCTGTGCCCACTCTGTGGTCTTTAAAGAAACAAAGTCCAAAGTCACTT	2820
Db	2761	AATGTGCACATCTCTGTGCCCACTCTGTGGTCTTTAAAGAAACAAAGTCCAAAGTCACTT	2820
OY	2821	TTGATGTGAACAAGAAAGAAAGAAATGAAGAAAGATGAGTCTAATATCAGCCTTAC	2880
Db	2821	TTGATGTGAACAAGAAAGAAAGAAATGAAGAAAGATGAGTCTAATATCAGCCTTAC	2880
OY	2881	AGACAGTTAATATATCTGTGCAGGCTTTTCTGTGTTGTCAGAAAGATTAAGCCAGTTGATA	2940
Db	2881	AGACAGTTAATATATCTGTGCAGGCTTTTCTGTGTTGTCAGAAAGATTAAGCCAGTTGATA	2940
OY	2941	ATGCCAATGTAGATCAAGAGGAGGCTGTAGGTTTGTCTATCATCTCACTTCAGAGGCA	3000
Db	2941	ATGCCAATGTAGATCAAGAGGAGGCTGTAGGTTTGTCTATCATCTCACTTCAGAGGCA	3000
OY	3001	ACGAAGTGCAGTCTTACTCCAAATTAACATGAGCTTTTACAAAACCCATATCGTATAC	3060
Db	3001	ACGAAGTGCAGTCTTACTCCAAATTAACATGAGCTTTTACAAAACCCATATCGTATAC	3060
OY	3061	CACCACTTTTCCCATGAGTCAATTTGTAAACCTAATGTAAAGAAAAATCTCTAGAGG	3120
Db	3061	CACCACTTTTCCCATGAGTCAATTTGTAAACCTAATGTAAAGAAAAATCTCTAGAGG	3120
OY	3121	AAAACTTTGAGGAACATTCATGTCACTGACCTGAAAGAGAAATGGGAAATGAGAACATTCCAA	3180
Db	3121	AAAACTTTGAGGAACATTCATGTCACTGACCTGAAAGAGAAATGGGAAATGAGAACATTCCAA	3180
OY	3181	GTACAGTGAGACACATTAAGCCGTATATACATTTAGAGAAATGTTTTTAAAGAACCAAGCT	3240
Db	3181	GTACAGTGAGACACATTAAGCCGTATATACATTTAGAGAAATGTTTTTAAAGAACCAAGCT	3240
OY	3241	CAACCAATATTAATGAGTAGTGGTCCAGTCTCTAATGAAGGGGCTCCAGATTAATGAATA	3300
Db	3241	CAACCAATATTAATGAGTAGTGGTCCAGTCTCTAATGAAGGGGCTCCAGATTAATGAATA	3300
OY	3301	TAGGTTCCAGTGATGAAGAACATTCAGACAGAACTAGGTAGAAAACAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTGATGAAGAACATTCAGACAGAACTAGGTAGAAAACAGAGGGCCAAATTTGA	3360

[illegible]

```

.QY      5521 GCACAGGGTGTCCACCCAAATTGGGTTGGTCAGACGAGATGCGTGGACAGAGACAATGGCT 5580
Db       5521 GCACAGGGTGTCCACCCAAATTGGTGTGTGCAGCCAGATGCTGTGCACAGAGACAATGGCT 5580
.QY      5581 TCCATGCAATTGGCGCAGATGTGTGAGGCACTGTGTGATGCCCGAGAGTGGTGTGGACA 5640
Db       5581 TCCATGCAATTGGCGCAGATGTGTGAGGCACTGTGTGATGCCCGAGAGTGGTGTGGACA 5640
.QY      5641 GGTGTGCACTTACCAGTGGCCAGAGCGTGGACACCTACTATATCCCAAGATGCCCCACA 5700
Db       5641 GGTGTGCACTTACCAGTGGCCAGAGCGTGGACACCTACTATATCCCAAGATGCCCCACA 5700
.QY      5701 GCCACTACTGA 5711
Db       5701 GCCACTACTGA 5711

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 16:19:58 ; Search time 7204.45 Seconds  
(without alignments)  
12838.232 Million cell updates/sec

Title: US-09-734-672-3  
Perfect score: 5711  
Sequence: 1 AGCTCCGTGAGACTTCTGTG.....TCCCCACAGACCACTACTGA 5711

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:  
2: em\_esthum:  
3: em\_estin:  
4: em\_estnu:  
5: em\_estov:  
6: em\_estpl:  
7: em\_estro:  
8: em\_hic:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hic:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: gb\_gss:  
18: em\_gss\_hum:  
19: em\_gss\_inv:  
20: em\_gss\_pln:  
21: em\_gss\_vrt:  
22: em\_gss\_fun:  
23: em\_gss\_mam:  
24: em\_gss\_mus:  
25: em\_gss\_other:  
26: em\_gss\_pro:  
27: em\_gss\_rod:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2061.4	36.1	2101 11	BC012577 Homo sapi
2	2055.6	36.0	2090 11	BC030969 Homo sapi
3	861.4	15.1	962 14	BQ068830 AGENCOURT
4	843	14.8	1089 13	BM452288 AGENCOURT
5	811.8	14.2	845 9	AU122476 AU122476
6	761.8	13.3	783 9	AU142729 AU142729

7	740.4	13.0	743 12	BG681276	BG681276 602627125
8	734.6	12.9	899 14	BQ422380	AGENCOURT
9	706.2	12.4	878 14	BQ215100	AGENCOURT
10	688	12.0	702 9	AU125312	AU125312 AU125312
11	687.6	12.0	747 12	BG777447	BG777447 602664758
12	686.6	12.0	739 12	BF508987	BF508987 UI-H-B14-
13	686.4	12.0	987 14	BM800251	AGENCOURT
14	657	11.5	900 12	BF791668	BF791668 602251822
15	651.6	11.4	938 12	BG257190	BG257190 602377909
16	639	11.2	675 9	AL704228	AL704228 DKFZ686K
17	619.2	10.8	921 12	BG178466	BG178466 602330271
18	603.8	10.6	630 9	A1992040	A1992040 ws44c12.x
19	593	10.4	906 14	BQ878445	BQ878445 AGENCOURT
20	593	10.4	909 14	BQ676829	BQ676829 AGENCOURT
21	593	10.4	933 14	BQ683955	BQ683955 AGENCOURT
22	592.6	10.4	800 12	BF794879	BF794879 602256742
23	591.4	10.4	627 9	A1589028	A1589028 C195d12.x
24	591	10.3	724 9	A1915085	A1915085 tq98f04.x
25	586	10.3	947 14	BQ679749	BQ679749 AGENCOURT
26	581	10.2	910 14	BQ677666	BQ677666 AGENCOURT
27	575	10.1	602 13	BE043993	BE043993 h83c01.x
28	548.2	9.6	638 13	BM042282	BM042282 603616359
29	536.8	9.4	638 10	BE264293	BE264293 601190118
30	531.6	9.3	563 10	AW295197	AW295197 UI-H-B12-
31	530.2	9.3	585 10	AW968546	AW968546 EST380622
32	518	9.1	666 10	AW968720	AW968720 EST380796
33	499.8	8.8	509 10	AW504244	AW504244 UI-HF-BNO
34	492.8	8.6	518 9	AA804632	AA804632 ob98d04.s
35	491.8	8.6	523 9	AA702344	AA702344 z187e03.s
36	491.2	8.6	509 10	BE018878	BE018878 db22c09.y
37	479.4	8.4	1220 14	BQ214737	BQ214737 AGENCOURT
38	472.2	8.3	487 9	AA812019	AA812019 ob39g02.s
39	450.6	7.9	575 10	BE564528	BE564528 601343863
40	449.8	7.9	453 9	AT684595	AT684595 wa84d07.x
41	444	7.8	444 9	AA486004	AA486004 ab38d03.r
42	437.4	7.7	488 9	AA481498	AA481498 oc07905.s
43	421.8	7.4	429 9	AA484941	AA484941 aa41a08.r
44	420.6	7.4	459 9	AT040685	AT040685 ox18b06.s
45	418.2	7.3	443 14	BQ308670	BQ308670 MKO-BT600

## ALIGNMENTS

RESULT 1  
LOCUS BC012577 2101 bp mRNA linear HTC 20-AUG-2001  
DEFINITION Homo sapiens, clone IMAGE:3996658, mRNA.  
ACCESSION BC012577  
VERSION BC012577.1 GI:15214876  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2101)  
TITLE Strausberg, R.  
JOURNAL Direct Submission  
Submitted (15-AUG-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-femail.nih.gov](mailto:cgaps-femail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 19 Row: a Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA 91: 6552298  
 This clone has the following problem: incomplete processing.

## FEATURES

source

1. 2101

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="399658"

/tissue\_type="Bladder, carcinoma"

/clone\_lib="NIH\_MGC\_53"

/lab\_host="DH10B"

/note="Vector: pDNR-LIB"

BASE COUNT 756 a 385 c 453 g 508 t  
 ORIGIN

## Query Match

Best Local Similarity 99.4%; Pred. No. 0;

Matches 2079; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```

1  AGCTGGCTGAGACTTCTGAGACCCGACACGAGCTGTGGGTTTCTCAGATAACTGGCC 60
6  AGCTGGCTGAGACTTCTGAGACGGGGA-CAGGCTGGGTTTCTCAGATACTGGCC 64
61  CCTGGCTGAGAGGCTTACCCCTGCTGGGTTAAAGTTCAATGGAACAGAAAGAA 120
65  CCTGGCTGAGAGGCTTACCCCTGCTGGGTTAAAGTTCAATGGAACAGAAAGAA 124
121  TGGATTATCTGCTCTTCCGCTTGAAGAAGTACAAATGTCATTAATGCTATGACAGAAA 180
125  TGGATTATCTGCTCTTCCGCTTGAAGAAGTACAAATGTCATTAATGCTATGACAGAAA 184
131  TCTTGAAGTCCCATCTGCTGAGTTGATCAAGAACCTGCTCCAAAGTGTGACC 240
135  TCTTGAAGTCCCATCTGCTGAGTTGATCAAGAACCTGCTCCAAAGTGTGACC 244
241  ACATATTTTGGCAATTTTGGCATGCTGAACCTTCTCAACCAAGAAAGAGGCTTTCACAGT 300
245  ACATATTTTGGCAATTTTGGCATGCTGAACCTTCTCAACCAAGAAAGAGGCTTTCACAGT 304
301  GTCTCTTATGTAAGATGATATACCAAAAGAGCTTACAGAAAGTACGAGATTATGTC 360
305  GTCTCTTATGTAAGATGATATACCAAAAGAGCTTACAGAAAGTACGAGATTATGTC 364
361  AACTGTTGAAGAGTATGAAATCATTTGCTTTTTCAGCTGTGACAGAGTTGGAGT 420
365  AACTGTTGAAGAGTATGAAATCATTTGCTTTTTCAGCTGTGACAGAGTTGGAGT 424
421  ATGCAACAGCTATATTTTGGCAAAAGAAATTAATCTCTCTGAACTCTTAAAGATG 480
425  ATGCAACAGCTATATTTTGGCAAAAGAAATTAATCTCTCTGAACTCTTAAAGATG 484
481  AAGTTTCTATCTCAAAAGTATGGCTTACAGAAACCGCTCCAAAGAGCTTTTACAGAGTG 540
485  AAGTTTCTATCTCAAAAGTATGGCTTACAGAAACCGCTCCAAAGAGCTTTTACAGAGTG 544
541  AACCCGAATCTCTCTGAGAAACAGTCTCAGTCCCAACCTCTTAACCTTGGAA 600
545  AACCCGAATCTCTCTGAGAAACAGTCTCAGTCCCAACCTCTTAACCTTGGAA 604
601  CTGTGAGAACTCTGAGAAACAGGATACAACTCTCAAAAGAGCTTGTCTACATTG 660
605  CTGTGAGAACTCTGAGAAACAGGATACAACTCTCAAAAGAGCTTGTCTACATTG 664
661  AATGGGATCTGATCTCTGAAAGTACCGTTAATTAAGCACTTATTCAGATGTGGAG 720
665  AATGGGATCTGATCTCTGAAAGTACCGTTAATTAAGCACTTATTCAGATGTGGAG 724
721  ATCAAGATCTGTAACCAATCCCTCAAGAACAGGATGAATCACTTGGATTGCG 780
|||||

```

```

725  ATCAAGATCTGTAACCAATCCCTCAAGAACAGGATGAATCACTTGGATTGCTG 784
781  CAAAAAGGCTGCTGTAATTTTTCAGAGAGATGTAACAAATCTCAATCATCTAC 840
785  CAAAAAGGCTGCTGTAATTTTTCAGAGAGATGTAACAAATCTCAATCATCTAC 844
841  CCAGTATATGATTTTGAACACCACTGAGAACGCTGACAGCTGAGAGGATCCAGAAAGT 900
845  CCAGTATATGATTTTGAACACCACTGAGAACGCTGAGAGGATCCAGAAAGT 904
901  ATCAGGATAGTCTGTTTCAACCTGATGTGAGACCATGTGGCAAAATCTCATGCA 960
905  ATCAGGATAGTCTGTTTCAACCTGATGTGAGACCATGTGGCAAAATCTCATGCA 964
961  GCTCATTAAGCATGAGAACACAGTCTTCTACTCACTAAGACAGTAATGATGAGAA 1020
965  GCTCATTAAGCATGAGAACACAGTCTTCTACTCACTAAGACAGTAATGATGAGAA 1024
1021  AGGCTGAATTTCTGTAATAAAGCAACAGCCTGCTTGAAGAGACCAACATACAGAT 1080
1025  AGGCTGAATTTCTGTAATAAAGCAACAGCCTGCTTGAAGAGACCAACATACAGAT 1084
1081  GGGCTGGAAGTAAAGAACATGTAATGATAGCGGACTCCACACAGAAAAAGTAG 1140
1085  GGGCTGGAAGTAAAGAACATGTAATGATAGCGGACTCCACACAGAAAAAGTAG 1144
1141  ATCTGAATCTGATCCCTGCTGAGAGAAAAAGTAATGAGAGAACTGTCATGCT 1200
1145  ATCTGAATCTGATCCCTGCTGAGAGAAAAAGTAATGAGAGAACTGTCATGCT 1204
1201  CAGAGAACTCTGAGATACGTAAGATGTTCTTGGATTAACCTAATAGCAGATTGAGA 1260
1205  CAGAGAACTCTGAGATACGTAAGATGTTCTTGGATTAACCTAATAGCAGATTGAGA 1264
1261  AAGTTAATGAGTGTGTTTCCAGAACTGTAAGCTTGTGATGATGATGATGATGATG 1320
1265  AAGTTAATGAGTGTGTTTCCAGAACTGTAAGCTTGTGATGATGATGATGATGATG 1324
1321  GGGAGTCTGAATCAAAATGCAAGAGTGTGATGTTTGGAGCTTCTAATAGGATAGT 1380
1325  GGGAGTCTGAATCAAAATGCAAGAGTGTGATGTTTGGAGCTTCTAATAGGATAGT 1384
1381  AATATTTCTGTTCTTCAAGAAATAGACTTACGTGCGAGTATCTCATGAGCTTTAA 1440
1385  AATATTTCTGTTCTTCAAGAAATAGACTTACGTGCGAGTATCTCATGAGCTTTAA 1444
1441  TATGTAAGTGAAGAGTGTCTCCCAATCATGTAAGATGAATTAAGCAAAATAT 1500
1445  TATGTAAGTGAAGAGTGTCTCCCAATCATGTAAGATGAATTAAGCAAAATAT 1504
1501  TTGGGAAAACTTATCGAAGAAAGGCAAGCTTCCCACTTAAGCATGTAAGCAAAATC 1560
1505  TTGGGAAAACTTATCGAAGAAAGGCAAGCTTCCCACTTAAGCATGTAAGCAAAATC 1564
1561  TAATTAATGAGCAATTTTGTACTGAGCCACAGATTAATCAAGAGCTTCCCTCAAAATA 1620
1565  TAATTAATGAGCAATTTTGTACTGAGCCACAGATTAATCAAGAGCTTCCCTCAAAATA 1624
1621  AATTAAGGCTTAAAGAGACCTTACATCAGGCTTCACTGAGAGATTTTATCAAGAAAG 1680
1625  AATTAAGGCTTAAAGAGACCTTACATCAGGCTTCACTGAGAGATTTTATCAAGAAAG 1684
1681  CAGATTTGGCACTTCAAAAGACTCTCTGAATGATTAATCAAGGAGTAACCAAAAGGAGC 1740
1685  CAGATTTGGCACTTCAAAAGACTCTCTGAATGATTAATCAAGGAGTAACCAAAAGGAGC 1744
1741  AGAATGTCAGATGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATG 1800
1745  AGAATGTCAGATGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATG 1804
1801  CTATTCAAGATGAGAAAAATCTTAACCAATGATGATGATGATGATGATGATGATGATG 1860
1805  CTATTCAAGATGAGAAAAATCTTAACCAATGATGATGATGATGATGATGATGATGATG 1864

```

QY 1861 AACGAAAGCTGACCTATTAAGCAGCAGTATAAGCAATATGGAATTCGATTAATATATCC 1920  
DB 1865 AACGAAAGCTGACCTATTAAGCAGCAGTATAAGCAATATGGAATTCGATTAATATATCC 1924  
QY 1921 ACAATTCAAAAGCACCATAAAGAAATAGGCTGAGAGAGAAATCTTCTACAGCATATTC 1980  
DB 1925 ACAATTCAAAAGCACCATAAAGAAATAGGCTGAGAGAGAAATCTTCTACAGCATATTC 1984  
QY 1981 ATGCGCTGACCTAGTACAGTACAGTAAATCTAAGCCCACTAATTTACTGAATTCGAA 2040  
DB 1985 ATGCGCTGACCTAGTACAGTACAGTAAATCTAAGCCCACTAATTTACTGAATTCGAA 2044  
QY 2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACACCAAA 2091  
DB 2045 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACACCAAA 2095

RESULT 2  
BC030969 2090 bp mRNA linear HNC 13-JUN-2002  
LOCUS Homo sapiens, Similar to breast cancer 1, early onset, clone  
DEFINITION IMAGE:4804551, mRNA:  
ACCESSION BC030969  
VERSION BC030969.1 GI:21411299  
KEYWORDS HTC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2090)  
AUTHORS Strausberg, R.  
TITLE Direct Submission.  
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: DCTD/DP  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdickpaxil.stanford.edu](mailto:mdickpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRL Plate: 41 Row: p Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA g1: 6552298  
This clone has the following problem: frame shifted.  
FEATURES  
source  
1..2090  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4804551"  
/issue\_type="Prostate, adenocarcinoma."  
/clone\_lib="NIH\_MGC\_60"  
/lab\_host="DH10B"  
/note="vector: pDNR-LIB"  
BASE COUNT 754 a 385 c 444 g 507 t  
ORIGIN  
Query Match 36.0%; Score 2055.6; DB 11; Length 2090;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2075; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
QY 1 AGCTGCGTAGACCTTCTGAGACCCGACACAGCTGTGGGGTTTCTCAGATAACTGGGCC 60

DB 2 AGCTGCGTAGACCTTCTGAGACCCGACACAGCTGTGGGGTTTCTCAGATAACTGGGCC 60  
QY 61 CCTGCGCTGAGAGGCGCTTCACCTCTGCTCTGGGTAAGTTATATTGGAACAGAAAGAA 120  
DB 61 CCTGCGCTGAGAGGCGCTTCACCTCTGCTCTGGGTAAGTTATATTGGAACAGAAAGAA 120  
QY 121 TGATATTATCTGCTCTGCGTTGAAGAGTACAAATGTCATTAAATGCTATGCAGAAA 180  
DB 121 TGATATTATCTGCTCTGCGTTGAAGAGTACAAATGTCATTAAATGCTATGCAGAAA 180  
QY 181 TCTTAGAGTGTCCCATCTGCTGAGTTGATCAAGAAACCTGTCTCACAAGCTGTACC 240  
DB 181 TCTTAGAGTGTCCCATCTGCTGAGTTGATCAAGAAACCTGTCTCACAAGCTGTACC 240  
QY 241 ACATATTTTGCAAAATTTTGCGATGCTTAACCTGTCAACCAAGAAAGAGGCTTCACAGT 300  
DB 241 ACATATTTTGCAAAATTTTGCGATGCTTAACCTGTCAACCAAGAAAGAGGCTTCACAGT 300  
QY 301 GTCCCTTATGTAGAATGATATTAACCAAGAGGCTTACAGAAAGTACGATTTAGTC 360  
DB 301 GTCCCTTATGTAGAATGATATTAACCAAGAGGCTTACAGAAAGTACGATTTAGTC 360  
QY 361 AACCTGTGAAGAGCTATTGAANAATCATTGCTTTGAGCTTGAACAGGTTTGAGT 420  
DB 361 AACCTGTGAAGAGCTATTGAANAATCATTGCTTTGAGCTTGAACAGGTTTGAGT 420  
QY 421 ATGCAAAACGCTATATTTTGCAAAAAGAAATTAACCTCTGTAACATCTTAAAGATG 480  
DB 421 ATGCAAAACGCTATATTTTGCAAAAAGAAATTAACCTCTGTAACATCTTAAAGATG 480  
QY 481 AAGTTCTATCATCCAAAGATAGGCTTACAGAAACCGTGCAAAAGACTTCTACAGATG 540  
DB 481 AAGTTCTATCATCCAAAGATAGGCTTACAGAAACCGTGCAAAAGACTTCTACAGATG 540  
QY 541 AACCCGAAATATCCTTCTGCAAGAAACCAAGTCTGAGTCAACTCTTAACCTTGAA 600  
DB 541 AACCCGAAATATCCTTCTGCAAGAAACCAAGTCTGAGTCAACTCTTAACCTTGAA 600  
QY 601 CTTGTGAACCTGAGAGCAAAAGCAGGATACACCTCAAAAGAGCTGTCTACATTTG 660  
DB 601 CTTGTGAACCTGAGAGCAAAAGCAGGATACACCTCAAAAGAGCTGTCTACATTTG 660  
QY 661 AATTGGATCTGATTTCTTGTGAAGATACCGTTATTAAGCAACTTTCAGTGTGGAG 720  
DB 661 AATTGGATCTGATTTCTTGTGAAGATACCGTTATTAAGCAACTTTCAGTGTGGAG 720  
QY 721 ATCAGAAATTTGTACAAATCACCCTCAAGGAACCGAGATGAATCAGTTTGGATTCTG 780  
DB 721 ATCAGAAATTTGTACAAATCACCCTCAAGGAACCGAGATGAATCAGTTTGGATTCTG 780  
QY 781 CAAAAAAGGCTGTTGAAATTTTTCGACAGGATTAACAAATACTGAAACATCATCAAC 840  
DB 781 CAAAAAAGGCTGTTGAAATTTTTCGACAGGATTAACAAATACTGAAACATCATCAAC 840  
QY 841 CCAGTATATATGATTTTGAACACACCTGAGAGGCTGAGAGGCAATCCGAAAAGT 900  
DB 841 CCAGTATATATGATTTTGAACACACCTGAGAGGCTGAGAGGCAATCCGAAAAGT 900  
QY 901 ATCAGGATGTTCTGTTCAAACTTGATGTGAGGCAATGTGGCACAATACTCATGCCA 960  
DB 901 ATCAGGATGTTCTGTTCAAACTTGATGTGAGGCAATGTGGCACAATACTCATGCCA 960  
QY 961 GCTCATTTACAGCATGAGACAGCAGTTTATCTACTCTAAAGCAGATGATGTGAAA 1020  
DB 961 GCTCATTTACAGCATGAGACAGCAGTTTATCTACTCTAAAGCAGATGATGTGAAA 1020  
QY 1021 AGGCTGAATTTCTGTAATAAAGCAACAGCCTGAGCTTACAGAGAGCAACATTAACAGAT 1080  
DB 1021 AGGCTGAATTTCTGTAATAAAGCAACAGCCTGAGCTTACAGAGAGCAACATTAACAGAT 1080  
QY 1081 GGGCTGGAAGTAAGCAACATGTAATGATAGCGGACTCCAGCACAGAAAAAAGGTAG 1140  
DB 1081 GGGCTGGAAGTAAGCAACATGTAATGATAGCGGACTCCAGCACAGAAAAAAGGTAG 1140

```

Db      1081 GGGCTGAAGTAAGCAACATGTAAATGATAGCGGACCTCCACACAGAAAAAAGTAG 1140
Qy      1141 ATCTGAATGCTATCCCTGTGTGAGAGAAAGTAATAGCAAACTGCCATGCT 1200
Db      1141 ATCTGAATGCTATCCCTGTGTGAGAGAAAGTAATAGCAAACTGCCATGCT 1200
Qy      1201 CAGAGATCTCTAGAGTACTGAGATGTTCTTGGATTAACACTAAATACAGCATTCACA 1260
Db      1201 CAGAGATCTCTAGAGTACTGAGATGTTCTTGGATTAACACTAAATACAGCATTCACA 1260
Qy      1261 AAGTTAATGAGTGGTTTCCAGAGTGAAGTGTAGGTTCTGATGACTCATGATG 1320
Db      1261 AAGTTAATGAGTGGTTTCCAGAGTGAAGTGTAGGTTCTGATGACTCATGATG 1320
Qy      1321 GGGAGTCTGAATCAATGCCAAGTAGTGTATGTAGCGTCTTAATAGGTTGATG 1380
Db      1321 GGGAGTCTGAATCAATGCCAAGTAGTGTATGTAGCGTCTTAATAGGTTGATG 1380
Qy      1381 AATATTTCTGTTCTTCAGAGAAATAGACTTAAGTGTAGGTTCTTAA 1440
Db      1381 AATATTTCTGTTCTTCAGAGAAATAGACTTAAGTGTAGGTTCTTAA 1440
Qy      1441 TATGTAAAGTGAAGAGTCTCTCAATCAGTGAAGTATATTTGAAGACAAATAT 1500
Db      1441 TATGTAAAGTGAAGAGTCTCTCAATCAGTGAAGTATATTTGAAGACAAATAT 1500
Qy      1501 TTGGGAAACCTATGGAAGGCAAGGCTCCCACTTAAGCCATGTAACGAAATAT 1560
Db      1501 TTGGGAAACCTATGGAAGGCAAGGCTCCCACTTAAGCCATGTAACGAAATAT 1560
Qy      1561 TAAATATGAGACATTTGTTACTGAGCCACAGATATATACAAAGCCTCCCTCACAAATA 1620
Db      1561 TAAATATGAGACATTTGTTACTGAGCCACAGATATATACAAAGCCTCCCTCACAAATA 1620
Qy      1621 AATTAACGCTAAAGGAGACCTACATCAGGCTTCATCTGAGATTTTATACAAAG 1680
Db      1621 AATTAACGCTAAAGGAGACCTACATCAGGCTTCATCTGAGATTTTATACAAAG 1680
Qy      1681 CAGATTGGCAGTTCAGAAAGACTCCTGAATGATTAATCAGGGAACCTAACCAAGCGAGC 1740
Db      1681 CAGATTGGCAGTTCAGAAAGACTCCTGAATGATTAATCAGGGAACCTAACCAAGCGAGC 1740
Qy      1741 AGAATGTCAGTGAATATTTACTAATAGTGTATGAGATTAATAAAGGCTGAT 1800
Db      1741 AGAATGTCAGTGAATATTTACTAATAGTGTATGAGATTAATAAAGGCTGAT 1800
Qy      1801 CTATTCAGAAATGAGAAAAATCCTAACCCATAGATCACTCGAAAAAGATCTGCTTGA 1860
Db      1801 CTATTCAGAAATGAGAAAAATCCTAACCCATAGATCACTCGAAAAAGATCTGCTTGA 1860
Qy      1861 AAACGAAAGCTGAACCTATTAACGACAGATATAGCAATATGGAACCTCGAATTAATATCC 1920
Db      1861 AAACGAAAGCTGAACCTATTAACGACAGATATAGCAATATGGAACCTCGAATTAATATCC 1920
Qy      1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAGTCTTACAGGATATTC 1980
Db      1921 ACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAGTCTTACAGGATATTC 1980
Qy      1981 ATGCGCTGACTAGTACTAGTCACTAGAAATCTAAAGCCACTAATGTTACTGTAATTCGAAA 2040
Db      1981 ATGCGCTGACTAGTACTAGTCACTAGAAATCTAAAGCCACTAATGTTACTGTAATTCGAAA 2040
Qy      2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACA 2090
Db      2041 TTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACA 2090

```

```

RESULT 3
LOCUS   B0068830
DEFINITION AGENCOURT_6740238 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5802685
ACCESSION B0068830

```

```

VERSION B0068830.1 GI:19897888
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 962)
JOURNAL NIH-MGC http://mhc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2037 row: h column: 14
High quality sequence stop: 700.
FEATURES
source
Location/Qualifiers
1..962
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5802685"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GCCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 302 a 195 c 235 g 229 t 1 others
ORIGIN
Query Match 15.1%; Score 861.4; DB 14; Length 962;
Best Local Similarity 96.2%; Pred. No. 1.6e-192;
Matches 890; Conservative 0; Mismatches 33; Indels 2; Gaps 1;
Qy 4451 TCCAGAACCAACGATCAGAAAAAGCAGTATTACTCAGAAAAAGTAGTAATCCG 4510
Db 15 TTGGCAGACGACGACATCAGAAAAAGCAGTATTACTCAGAAAAAGTAGTAATCC 74
Qy 4511 TATAGCCGAAATCCAGAAAGGCTTTCGCTGACAAAGTTTGAGGTCTGCAGATATTC 4570
Db 75 TATAGCCGAAATCCAGAAAGGCTTTCGCTGACAAAGTTTGAGGTCTGCAGATATTC 134
Qy 4571 TACAGTAAATAAATAAGAACCAAGAGTGAAGAGTATCCCTTCTAATGCCATAT 4630
Db 135 TACAGTAAATAAATAAGAACCAAGAGTGAAGAGTATCCCTTCTAATGCCATAT 194
Qy 4631 AGATGATGAGTGTGATGACAGTGTGCTGGGAGTCTTGAATGAATACCATC 4690
Db 195 AGATGATGAGTGTGATGACAGTGTGCTGGGAGTCTTGAATGAATACCATC 254
Qy 4691 TCAAGAGAGCTCATTAAGTGTGATGAGTGAAGAGCAGACGCTGAAGAGTCTGGGCC 4750
Db 255 TCAAGAGAGCTCATTAAGTGTGATGAGTGAAGAGCAGACGCTGAAGAGTCTGGGCC 314
Qy 4751 ACACGATTTGACGGAACATCTTACTCCAAAGCAAGATTAAGAGGAAACCTTACT 4810
Db 315 ACACGATTTGACGGAACATCTTACTCCAAAGCAAGATTAAGAGGAAACCTTACT 374
Qy 4811 GGAATCTGGAATCAGCCTCTCTGATGACCCGTAATCTATCTCTGTAAGACAGAGC 4870
Db 375 GGAATCTGGAATCAGCCTCTCTGATGACCCGTAATCTATCTCTGTAAGACAGAGC 434

```

```

QY 4871 CCCAGATCAGCTCGTGTGGCAACATACCACCTTCACTTCACTTGAAGTTCGCCA 4930
    |||
Db 4435 CCCAGATCAGCTCGTGTGGCAACATACCACCTTCACTTGAAGTTCGCCA 4934
    |||
QY 4931 ATTGAAGTTGCAAGATCTGCCAGAGTCCAGTCTGCTCACTACTGATCTGCTGG 4990
    |||
Db 495 ATTGAAGTTGCAAGATCTGCCAGAGTCCAGTCTGCTCACTACTGATCTGCTGG 554
    |||
QY 4991 GTATATGCAATGGAAGATGTGAGCGGGGAGGAGCAGATTTGACAGCTTCAACAGA 5050
    |||
Db 555 TTATATGCAATGGAAGATGTGAGCGGGGAGGAGCAGATTTGACAGCTTCAACAGA 614
    |||
QY 5051 AAGGTCACAAAGAAGATGTCATGTGTGTCTGCTGACCCAGAGAAATTTATGCT 5110
    |||
Db 615 AAGGTCACAAAGAAGATGTCATGTGTGTCTGCTGACCCAGAGAAATTTATGCT 674
    |||
QY 5111 CGGTGCAAGTTGCCAGAAACACCACTACTTTAACTAATCTAATTTCTGAAGAGAC 5170
    |||
Db 675 CGGTGCAAGTTGCCAGAAACACCACTACTTTAACTAATCTAATTTCTGAAGAGAC 734
    |||
QY 5171 TACTATGTTGTATGAAACAGATGCTGAGTTGTGTGTGAGCGGACCTGAATATT 5230
    |||
Db 735 TACTATGTTGTATGAAACAGATGCTGAGTTGTGTGTGAGCGGACCTGAATATT 794
    |||
QY 5231 TCTAGCAATTTGCGGAGAGAAATGGTACTATTCTGAGGAGCCAGTCTATTAA 5290
    |||
Db 795 TCTAGCAATTTGCGGAGAGAAATGGTACTATTCTGAGGAGCCAGTCTATTAA 854
    |||
QY 5291 AGAAGAAAAATCTGAATGAGCATATTTTGAAGTCAGAGAGATGTG--TCAATGGA 5348
    |||
Db 855 AGAAGAAAAATCTGAATGAGCATATTTTGAAGTCAGAGAGATGTGCAATGGA 914
    |||
QY 5349 AGAAGCCACCAAGGTCCAAAGCGAG 5373
    |||
Db 915 AGAAGCCACCAAGGTCCAAAGCGAG 939
    |||

RESULT 4
BM452288 1089 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6386302 NIH_MGC_72 Homo sapiens cdna clone IMAGE:5356666
DEFINITION 5', mRNA sequence.
ACCESSION BM452288
VERSION BM452288.1 GI:18501328
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 1089)
          NIH-MGC http://mgi.nci.nih.gov/
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: ATCC/DCFD/DPF
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LLAM12200 row: k column: 19
          High quality sequence stop: 667.
          Location/Qualifiers
            1..1089
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:5356666"
              /clone_1lb="NIH_MGC_72"
              /tissue_type="melanotic melanoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: skin; Vector: pCMV-Sport6; site_1: Not;"

```

```

BASE COUNT      389 a      212 c      248 g      240 t
ORIGIN
Query Match      14.8%; Score 843; DB 13; Length 1089;
Best Local Similarity 96.1%; Pred. No. 3.9e-188;
Matches 929; Conservative 0; Mismatches 30; Indels 8; Gaps 6;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies.
QY 1925 TTCAAAAGCACCTTAAAGAAATAGGCTGAGAGAGAACTTCTTACCAGCATATTCATGC 1984
    |||
Db 13 TTCAAAAGCACCTTAAAGAAATAGGCTGAGAGAGAACTTCTTACCAGCATATTCATGC 72
    |||
QY 1985 GCTTGAACCTGTAGTCTGTAATATCAACCCCACTTAATGTACGAAATTTGA 2044
    |||
Db 73 GCTTGAACCTGTAGTCTGTAATATCAACCCCACTTAATGTACGAAATTTGA 132
    |||
QY 2045 TACTTGTCTAGCAGTCAGATGATTAAGAAAAAAGTACACCAATGCGCAGTCA 2104
    |||
Db 133 TACTTGTCTAGCAGTCAGATGATTAAGAAAAAAGTACACCAATGCGCAGTCA 192
    |||
QY 2105 CAGCAGAACTTCAACTCATGGAAGTAAAGAACTGCACTGAGAGCCAGAGACTAA 2164
    |||
Db 193 CAGCAGAACTTCAACTCATGGAAGTAAAGAACTGCACTGAGAGCCAGAGACTAA 252
    |||
QY 2165 CAAGCCAAATGACACAGCAAGTAAAGACATGACATGATCTTCCAGAGCTGAAGTT 2224
    |||
Db 253 CAAGCCAAATGACACAGCAAGTAAAGACATGACATGATCTTCCAGAGCTGAAGTT 312
    |||
QY 2225 AACAATGACACCTGCTTCTTCTTAAAGTAAATACAGTAAAGATTTGT 2284
    |||
Db 313 AACAATGACACCTGCTTCTTCTTAAAGTAAATACAGTAAAGATTTGT 372
    |||
QY 2285 CAATCTAGCTTCCAGAGAGAAAGAGAACTAGAAACAGTTAAAGTGTCTAA 2344
    |||
Db 373 CAATCTAGCTTCCAGAGAGAAAGAGAACTAGAAACAGTTAAAGTGTCTAA 432
    |||
QY 2345 TAATGCTGAAGACCCCAAGATCTCATGTAATGGAAGGTTTGAACCTGAAG 2404
    |||
Db 433 TAATGCTGAAGACCCCAAGATCTCATGTAATGGAAGGTTTGAACCTGAAG 492
    |||
QY 2405 ATCTGTAGAGATGAGATATTTTCATTTGCTGACTGATGATTAATGAGCACTGAGAAAG 2464
    |||
Db 493 ATCTGTAGAGATGAGATATTTTCATTTGCTGACTGATGATTAATGAGCACTGAGAAAG 552
    |||
QY 2465 TATCTGTTACTGGAAGTTAGCACTGAGGAGGCAAAACAGAACCAATTAATGTGT 2524
    |||
Db 553 TATCTGTTACTGGAAGTTAGCACTGAGGAGGCAAAACAGAACCAATTAATGTGT 612
    |||
QY 2525 GACTCAGTGTGACGATTTTGAAGAAACCCCAAGGACATTAATCATGTTTCCAAAGTAA 2584
    |||
Db 613 GACTCAGTGTGACGATTTTGAAGAAACCCCAAGGACATTAATCATGTTTCCAAAGTAA 672
    |||
QY 2585 TAAAGATGACACAGAGGCTTTAAGTATTCATTTGGGACATGAGATTACACAGTGGGA 2644
    |||
Db 673 TAAAGATGACACAGAGGCTTTAAGTATTCATTTGGGACATGAGATTACACAGTGGGA 732
    |||
QY 2645 AACAAGCATGAAGT--GGAAGAAAGTGAACCTGATGCTAGATTGTCAGAAATACATTC 2703
    |||
Db 733 AACAAGCATGAAGTGAAGAAAGTGAACCTGATGCTAGATTGTCAGAAATACATTC 792
    |||
QY 2704 AGTTTCAAGGCGCCAGTCTATTGCTCTGTTTCAATTCAGAGAAATGCGAAGAGAAAT 2763
    |||
Db 793 A-GTTTCAAGGCGCCAGTCTATTGCTCTGTTTCAATTC-GGAAGTGAAGAGAGAT 850
    |||
QY 2764 GTCAACATCTCTGCCACTCTGGGTGC-TTAAAGAAACAAAGTCCAAAGTCACTTTT 2822
    |||
Db 851 GTCAACATCTCTGGCCACTCTGGGCTTTTAAAGAAACAAAGTCCAAAGTCCCTTTT 910
    |||
QY 2823 GAATGTGAACAAA--GGAAGAAATTCAGAGAAAGATGATCTAA-TATCAAGCCTGT 2878
    |||
Db 911 GAATGTGGGACCAAAAGGACAAAATTCAGAGAAAGATGATGATCTAATATCAAGCCTTT 970
    |||

```

QY	2879	ACAGCA	2885	
Db	971	AACGAA	977	
RESULT 5				
LOCUS	AU122476			
DEFINITION	AU122476 MAMMAL Homo sapiens CDNA clone MAMMAL002447 5', mRNA	845 bp	mRNA	linear EST 01-AUG-2002
ACCESSION	AU122476			
VERSION	AU122476.1	GI:10937746		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 845) Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.			
TITLE	HRI human CDNA project			
JOORNAL	Unpublished (2000)			
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3875 Fax: 81-438-52-3986 Email: genomicehri.co.jp HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; CDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.			
FEATURES	Location/Qualifiers			
Source	1..845 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MAMMAL002447" /clone_lib="MAMMAL" /tissue_type="mammary gland" /note="Vector: pME18SFL"			
BASE COUNT	279 a 172 c 191 g 200 t	3 others		
ORIGIN				
Query Match	14.2%	Score 811.8;	DB 9;	Length 845;
Best Local Similarity	99.1%;	Prod. No. 8.3e-181;		
Matches 836;	Conservative 0;	Mismatches 5;	Indels 3;	Gaps 2;
QY	3819	GTAACAATATACCTTCTCAGTCTACTAGCATGACACCGTGGCTACGAGTGTCTGCT	3878	
Db	1	GTAACAATATACCTTCTCAGTCTACTAGCATGACACCGTGGCTACGAGTGTCTGCT	60.	
QY	3879	AAGAACACAGAGAGAAATTTATATCATTTGAAGAAATAGCTTAAATGACTGCAGTAACAG	3938	
Db	61	AAGAACACAGAGAGAAATTTATATCATTTGAAGAAATAGCTTAAATGACTGCAGTAACAG	120	
QY	3939	GTAATATTGGCAAAAGGACATCTCAGAACATACCTTGTGTGAGAAACAAATGTTCTGCT	3998	
Db	121	GTAATATTGGCAAAAGGACATCTCAGAACATACCTTGTGTGAGAAACAAATGTTCTGCT	180	
QY	3999	AGCTTGTCTTCTTACAGAGTGCAGTAATTTGGAAGTCTGACTGCAATATACAAACCCAG	4058	
Db	181	AGCTTGTCTTCTTACAGAGTGCAGTAATTTGGAAGTCTGACTGCAATATACAAACCCAG	240	
QY	4059	GATCTCTTCTTGTGATTTGTTCTTCCAAACAAATGAGCATCAGTCTGAAAGCCAGGAGTT	4118	
Db	241	GATCTCTTCTTGTGATTTGTTCTTCCAAACAAATGAGCATCAGTCTGAAAGCCAGGAGTT	300	
QY	4119	GGTGTGATGCAAGGAATGGCTTTCAGATGATGAACAAAGAGGAACGGGCTTGAGAA	4178	
Db	301	GGTGTGATGCAAGGAATGGCTTTCAGATGATGAACAAAGAGGAACGGGCTTGAGAA	360	

QY	4179	AATATATCAAGAGACCAAGCATGGATTCCTAAACCTTAAGTGAAGCAGCATCTGGCGTGTAG	4238
Db	361	AATATATCAAGAGACCAAGCATGGATTCCTAAACCTTAAGTGAAGCAGCATCTGGCGTGTAG	420
QY	4239	AGTGAACAAGCGCTCTGGAACACCTGCTCAGGGCTATTCCTCTCAGAGTACATTTTAAC	4298
Db	421	AGTGAACAAGCGCTCTGGAACACCTGCTCAGGGCTATTCCTCTCAGAGTACATTTTAAC	480
QY	4239	ACTCAGCAGAGGGATACCATGCAACATTAACCTGATTAAGCTCAGCAGGAATGGCTGAA	4358
Db	481	ACTCAGCAGAGGGATACCATGCAACATTAACCTGATTAAGCTCAGCAGGAATGGCTGAA	540
QY	4359	CTAGAAAGCTGTGTTTGAACACAGATGGGGCCAGCCCTTCAACAGCTACCCTTCCATCATTA	4418
Db	541	CTAGAAAGCTGTGTTTGAACACAGATGGGGCCAGCCCTTCAACAGCTACCCTTCCATCATTA	600
QY	4419	AGTGACTCTTCTGCCCCTTGAGAGACCTGGGAATCCAGAACCAAGCACATCCAGAAAAACA	4478
Db	601	AGTGACTCTTCTGCCCCTTGAGAGACCTGGGAATCCAGAACCAAGCACATCCAGAAAAACA	660
QY	4479	GTATTTAACCTTCACAGAAAAAGTAGTAATACCTATTAAGCCAGATCCAGAAAGCCCTTCT	4538
Db	661	GTATTTAACCTTCACAGAAAAAGTAGTAATACCTATTAAGCCAGATCCAGAAAGCCCTTCT	720
QY	4539	GCTGCACAGTTTGGAGCTGTGTCACATATGTTCTTACCACTAAATAATTAAGAACCCAGGA	4596
Db	721	GCTGCACAGTTTGGAGCTGTGTCACATATGTTCTTACCACTAAATAATTAAGAACCCAGGA	780
QY	4597	TGGAAGAAGTCA-TCGCCCTTCTTAATATGCCCATCATATTAGATGATAGGTGTACATGCACAGT	4655
Db	781	TGGAAGAAGTCAATTCCTTCTTAATATGCCCATCATATTAGATGATAGGTGTACATGCACAGT	840
QY	4656	TGCT 4659	
Db	841	TGCT 844	
RESULT 6	AU142729	783 bp	EST 05-AUG-2002
LOCUS	AU142729	Y79AA1 Homo sapiens cDNA clone Y79AA1000792 5', mRNA	
DEFINITION	AU142729	Y79AA1 Homo sapiens cDNA clone Y79AA1000792 5', mRNA	
ACCESSION	AU142729		
VERSION	AU142729		
KEYWORDS	AU142729.1	GI:11004250	
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 783) Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Iwaga,T.		
TITLE	HRI human cDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel.: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1. 783 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Y79AA1000792" /clone_id="Y79AA1"		



BASE COUNT	255 a	146 c	185 g	194 t	3 others
ORIGIN					

Query Match	13.3%	Score 761.8	DB 9	Length 783
Best Local Similarity	99.2%	Pred. No. 5.4e-169		
Matches 774, Conservative	0	Mismatches 5	Indels 1	Gaps 1

Qy	3579	TTTGATGATGTTGCTAAATTAAGAGAAATCTAGTTTGCGAAATGACATTTAAGAAAGT	3638
Db	1	TTTGATGATGCTAAATTAAGAGAAATCTAGTTTGCGAAATGACATTTAAGAAAGT	60

3639 TCTGCTGTTTTTATGCAAAAGCGTCCAGAAAGAGAGCTTAGCAGAGTCTTAGCCCTTC 3658  
 0Y  
 61 TCTGCTGTTTTTATGCAAAAGCGTCCAGAAAGAGAGCTTAGCAGAGTCTTAGCCCTTC 120  
 Db

0y 3699 ACCATTACACATTTGGCTCAGGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAA 3758  
 Db 121 ACCATACACATTTGGCTCAGGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAA 180

Dy 3759 GACAACTTATCTAGTGAGGATGAAGAGCTTCCTCGCTCCCAACTGTGTTATTTGGTAA 3818  
 Dy 3760 GACAACTTATCTAGTGAGGATGAAGAGCTTCCTCGCTCCCAACTGTGTTATTTGGTAA 3819  
 Db 181 GACAACTTATCTAGTGAGGATGAAGAGCTTCCTCGCTCCCAACTGTGTTATTTGGTAA 240

3819 GAAACAAATATACCTTCAGCTACCTAGGCAATAGCACCGTGTGCTGCT 3878  
 3819 |||||  
 241 GTAACCAATATACCTTCAGCTACTACTAGGCAATAGCACCGTGTGCTGCT 300

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
86

Db 361 GTATATTGGCAAGGCATCTCAGGACATCACTTAGTGAGCAACAATGTTCTGCT 420

Db 421 AGCTGTTTCTTCACAGTCAGTGAATTGGAAGACTTGCACCTGCAATATCAACACCCACG 480

Db - 481 GATCCTTCTGTGATGGTCTTCCAAACAATGAGGCATCAGTGTGAAGCCAGGAGTT 540

Db

541 GGCCTGATGACACAGGAATTGGTTTCAGTGAATGACAGAAAGAGGACGGGCTTGACAGA 600

[illegible]

4298 CATCAGAGAGAGGATACCATGGAACATATACCTTAAAGCTCCAGGACAGAAATGCGCTCA 4355  
661 AGTGAACAAGCGCTCTCTGGAAGACTGCTCAGGCGTATCCTCTCAGAGTACACATTTTAACC 720

Db 721 CACTCAGCAGANGATACCATGCNNCATTAACCTGATAAAGCTTCAGCAGAGAAATGGCTGA 780

LOCUS	DEFINITION	743 bp	mrna	linear	EST 01-MAY-2001
B6681276	6026677125F1 NCI CGAP Skn4 Homo sapiens cDNA clone IMAGE:4751887 5'				

Accession: BG6681276  
 Version: BG6681276.1  
 Keywords: EST.

ORGANISM Homo sapiens

REFERENCE  
Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Plimates: Carnivora: Homnidae: Homo.  
1 (phases 1 to 743)  
AUTHORS  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: James Cleaver, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

FEATURES	High quality sequence stop: 741.
SOURCE	Location/Qualifiers
	1. 743
	/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:4751887"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"

```

```

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo d
Average insert size 1.5kb. Library constructed by Biff

```

BASE COUNT	233 a	156 c	182 g	172 t
ORIGIN				

TECHNOLOGIES. NOTE: THIS IS A NCBI-CCAF LIBRARY.

Exact match	13.0%	Score 70-74	66 12%	Denglin 74%
Best Local Similarity	99.9%	Pred: No. 6e-164		
Matches 741; Conservative	0;	Mismatches 1;	Indels 0;	Gaps

Db 1 GGAGCAGCAGCTGGAAGAGTCTGGGCCACACAGATTGACGGAAACATCTTACTTGCCAG 6

Db 61 GCAGATCTAGAGGGAACCCCTTACCTGGATCTGGATCATCAGCCCTTCTCTGATGACCC 1

Db 121 TGAATCTGATCTTCTGAGACAGAGCCCCAGAGTCAGCTCGTTGGCAACATACCAATC 1

Db 181 TTCAACCTCTGATTTGAAGTTCCCAATTGAAGTTGCAATCTGCCAGAGTCCAGC 240

Db 241 TGCTGCTCATCTACTACTGCTGGGTATAATGCAATGGAGAAGTGTGACAGCGA 3

5084 TGGGCTGAAGCCCCAGAAATATTTATCGTCCTATCAAGCTTTGCCAGAAAAACACCAATCAC 5

Db 361 TGGCTGACCCGAGAGATTATGCTGCTGTACAGTTTGGCAGAAAACACCACATTCAC 41

Db 421 TTTACTAATCTAATTACTGAAGACACTCACTGTCGTATATGAACAACAGATGCTGACTT 4  
07 5204 TGTCGTGAACGCACACTGAAATATTTCTAGCAATTGCGGAGCAAAATGGGTAGCTTAG 5

Db	481	TGTGTGAAGCAGCACTGAATATTTTCTAGCAATTCCGGAGAGAAAATGGTACTTTAG	53
Qy	5264	CTATTTCTGGGTGACCCAGTCTATTAAAGAAGCAAAAATGCTGAAATGAGCATGATTTTGA	53

```

|||||
541 CTAATTTCTGGTGACCCAGCTATATTAAGAAAGAAAAAAGCTGAATGAGCATGATTTTGA 600
DB
5324 ACTCAGAGAGATGTCGTCATATGAGAAACCCAGCAGTCACCAAGCAGCAGAGAAATC 5383
QY
601 ACTCAGAGAGATGTCGTCATATGAGAAAGCAAGTCACCAAGCAGCAGAGAAATC 660
DB
5384 CCAGGACAGAAAGATCTTCAGGGGGCTAGAAATCTGTCATGAGCCCTTCACCAACAT 5443
QY
661 CCAGGACAGAAAGATCTTCAGGGGGCTAGAAATCTGTCATGAGCCCTTCACCAACAT 720
DB
5444 GCCCAGCATCTCACTGGAAATGC 5465
QY
721 GCCCAGCATCTCACTGGAAATGC 742
DB

RESULT 8
LOCUS B0422380 899 bp mRNA linear EST 23-MAY-2002
DEFINITION AGENCOURT_7802085 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042052
ACCESSION B0422380
VERSION B0422380.1 GI:21117695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 899)
NIH-MGC http://mgi.mcl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@brr-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13280 row: n column: 05
High quality sequence stop: 597.
Location/Qualifiers
1..899
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6042052"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 291 a 182 c 192 g 230 t 4 others
ORIGIN
Query Match 12.9%; Score 734.6; DB 14; Length 899;
Best Local Similarity 97.1%; Pred. No. 1.5e-162;
Matches 824; Conservative 0; Mismatches 11; Indels 14; Gaps 7;

```

```

DB
130 GTACAAATATGCTATTAATGCTATGCAAGAAATCTTAGAGTGTCCATCTGTGAGTTG 189
QY
210 ATCAAGCAACCTGTCTCCCAAGAGTGTCACCATATTTTGCATTTTGCATGCTGAA 269
DB
190 ATCAAGCAACCTGTCTCCCAAGAGTGTCACCATATTTTGCATTTTGCATGCTGAA 249
QY
270 CTTCACCAAGCAAGAAAGGGCTTCACAGTGTCTTTATGTAAAGATATTAACCA 329
DB
250 CTTCACCAAGCAAGAAAGGGCTTCACAGTGTCTTTATGTAAAGATATTAACCA 309
QY
330 AGAGCCTTCAAGAAAGTCAAGATTTAGTCACTTTTGAAGAGCTTTGAATATCTT 389
DB
310 AGAGCCTTCAAGAAAGTCAAGATTTAGTCACTTTTGAAGAGCTTTGAATATCTT 369
QY
390 TGTGCTTTTACAGTTCACAGAGTTTGGAGTATGCAACAGCTATATTTTGAAGAA 449
DB
370 TGTGCTTTTACAGTTCACAGAGTTTGGAGTATGCAACAGCTATATTTTGAAGAA 429
QY
450 GAAATATCTCTCTGACATCTAAAGATGAGTTTCTATCAGCAAGATATGAGGCTAC 509
DB
430 GAAATATCTCTCTGACATCTAAAGATGAGTTTCTATCAGCAAGATATGAGGCTAC 489
QY
510 AGAAGCCTGCAAGAAAGTCTTACAGAGTGAACCCGAAATCTCTCTGAGAAAC 569
DB
490 AGAAGCCTGCAAGAAAGTCTTACAGAGTGAACCCGAAATCTCTCTGAGAAAC 546
QY
570 AGTCTAGTGTCCACTCTCTACCTTGAACCTGGAACCTGGAACCAAGCAGG 629
DB
547 AGTCTAGTGTCCACTCTCTACCTTGAACCTGGAACCTGGAACCAAGCAGG 606
QY
630 ATACAACTCTCAAGAGCTCTCTACATTTGATTTGATTTCTTGAAGATAC 689
DB
607 ATACAACTCTCAAGAGCTCTCTACATTTGATTTGATTTCTTGAAGATAC 666
QY
690 GTTAATTAAGCAACTTATTGCAAGTGTGAGATGATGATTTCTTGAAGATAC 749
DB
667 GTTAATTAAGCAACTTATTGCAAGTGTGAGATGATGATTTCTTGAAGATAC 726
QY
750 GGACCA-GGAGTGAATAGTTGATTTCTCAAGAAAGCTG-TTGAATTTCTG 807
DB
727 GGACCAAGGATGATGATTTGATTTCTCAAGAAAGCTG-TTGAATTTCTG 786
QY
808 AGACGATGTAACAA-ATACGATGATC-ATCAACCCAGTATATGA-TTTGAACACA 864
DB
787 AGACGATGTAACAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 846
QY
865 CTGAGAAC 873
DB
847 CTGAGAAC 855

RESULT 9
LOCUS B0215100 878 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7591049 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6065516
ACCESSION B0215100
VERSION B0215100.1 GI:20396500
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 878)
NIH-MGC http://mgi.mcl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@brr-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```



```

Db      301 TGAACAGAGCTCTGAGAGACTGCTAGGGCTATCTCTCAGAGTGACATTTAAACAC 360
QY      4301 TCAGCAGAGGAGATACATGCAACATACGATTAAGCCTCCACAGGAATGGCTGACT 4360
Db      361 TCAGCAGAGGAGATACATGCAACATACGATTAAGCCTCCACAGGAATGGCTGACT 420
QY      4361 AGAAGCTGTGTAGAACAGCATGGAGCCAGCCTTCTACAGCTACCTTCATCATTAAG 4420
Db      421 AGAAGCTGTGTAGAACAGCATGGAGCCAGCCTTCTACAGCTACCTTCATCATTAAG 480
QY      4421 TGACCTCTTGTGCTGAGAGCCTGCGAAATCCAGAACAAAGACATCAGAAAAAGCAGT 4480
Db      481 TGACCTCTTGTGCTGAGAGCCTGCGAAATCCAGAACAAAGACATCAGAAAAAGCAGT 540
QY      4481 ATTAACTTCACAGAAAGTAGTAATACCTATAGCAGAAATCCAGAAAGCCTTCTGCG 4540
Db      541 ATTAACTTCACAGAAAGTAGTAATACCTATAGCAGAAATCCAGAAAGCCTTCTGCG 600
QY      4541 TGACAGATTGTGAGTGCTGACAGATAGTTCTACAGTAAATAAAGAACAGAGTGA 4600
Db      601 TGACAGATTGTGAGTGCTGACAGATAGTTCTACAGTAAATAAAGAACAGAGTGA 660
QY      4601 AAGGTCATCCCTCTTAATGCCCATCATTAAGTATAGTGG 4643
Db      661 AANGCAT-CCCTTCTAAATGCCCATCATTAAGTATAGTGG 702

```

## RESULT 11

```

Bg777447 747 bp mRNA linear EST 15-MAY-2001
LOCUS     602664758f1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:480451 5'
DEFINITION mRNA sequence.
Bg777447
ACCESSION Bg777447.1 GI:14047764
VERSION    EST
KEYWORDS   human.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE     1 (bases 1 to 747)
JOURNAL   NIH-MGC http://mgi.nci.nih.gov/
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-remail.nih.gov
          Tissue Procurement: DCTD/DRP
          cDNA Library Preparation: CLONTECH Laboratories, Inc.
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L10M1651 row: c column: 16
          High quality sequence stop: 723.
          Location/Qualifiers
            1..747
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:480451"
              /clone_lib="NIH_MGC_60"
              /tissue_type="adenocarcinoma"
              /lab_host="DH10B (T1 phage-resistant)"
              /note="Organ: prostate; Vector: pDNF-LIB (Clontech);
              Site.1: Sfil (ggcgccctggcc); Site.2: Sfil (ggcgccctggcc)
              ; Double-stranded cDNA was prepared from cell line RNA.
              5' and 3' adaptors were used in cloning as follows: 5'
              adaptor sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor
              sequence: 5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)-BN-3'
              (where B = A, C, or G and N = A, C, G, or T). Average
              insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
              contained inserts by PCR. This library was enriched for
              full-length clones and was constructed by Clontech

```

## FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:480451"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNF-LIB (Clontech);
Site.1: Sfil (ggcgccctggcc); Site.2: Sfil (ggcgccctggcc)
; Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)-BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech

```

Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

Query Match 12.0%; Score 687.6; DB 12; Length 747;  
 Best Local Similarity 97.7%; Pred. No. 1.8e-151;  
 Matches 729; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

```

QY      1 AGCTGCTGAGACTCTGAGAGCCTCCAGCAGGCTGTGGGTTCTCGATATCTGGCC 60
Db      3 AGCTGCTGAGACTCTGAGAGCCTCCAGCAGGCTGTGGGTTCTCGATATCTGGCC 61
QY      61 CCGCCCTGAGAGGCTTCCACCTTCGCTGAGGTTGAAGTTCAATGGAACAGAAAGAA 120
Db      62 CCGCCCTGAGAGGCTTCCACCTTCGCTGAGGTTGAAGTTCAATGGAACAGAAAGAA 121
QY      121 TGATTTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGTCAATATGCTATGCAAGAA 180
Db      122 TGATTTATCTGCTCTTCCGCTTGAAGAAAGTACAAATGTCAATATGCTATGCAAGAA 181
QY      181 TCTTAGAGTGCTCCATCTGCTGAGTGTATGATGAGAACTGTGCTCCAAAGTGTGACC 240
Db      182 TCTTAGAGTGCTCCATCTGCTGAGTGTATGATGAGAACTGTGCTCCAAAGTGTGACC 241
QY      241 ACATATTTTGCAGAAATTTTGCAGTGTGAACCTTCTGACAGAGAAAGGCTTCCACAGT 300
Db      242 ACATATTTTGCAGAAATTTTGCAGTGTGAACCTTCTGACAGAGAAAGGCTTCCACAGT 301
QY      301 GTCTTTTATGTAAGAAATGATATACCAAGAGCCTTACAGAAAGTACAGATTAGTC 360
Db      302 GTCTTTTATGTAAGAAATGATATACCAAGAGCCTTACAGAAAGTACAGATTAGTC 361
QY      361 AACTGTGTGAAGAGCAATGAA-AATCATTTGCTTTACCTTGACCTGACAGATTGGAG 419
Db      362 AACTGTGTGAAGAGCAATGAA-AATCATTTGCTTTACCTTGACCTGACAGATTGGAG 421
QY      420 TATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCTGCAACATCTAAAGAT 479
Db      422 TATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCTGCAACATCTAAAGAT 481
QY      480 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTGCCAAAGACTTCTACAGAGT 539
Db      482 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTGCCAAAGACTTCTACAGAGT 541
QY      540 GAACCCGAAATTCCTCTGCGAGAAACAGCTGACGTCACCTGTAACCTTGA 599
Db      542 GAACCCGAAATTCCTCTGCGAGAAACAGCTGACGTCACCTGTAACCTTGA 601
QY      600 ACTGTGAGAACTCTGAGCAAGCAAGCGGATACAACTCAAAAGACGCTGTCTACATT 659
Db      602 ACTGTGAGAACTCTGAGCAAGCAAGCGGATACAACTCAAAAGACGCTGTCTACATT 661
QY      660 GAATTGGGATCGATTCTTCTGAAGATACCGTTAATAAGCACTTATTCAGAGT-666 718
Db      662 GAATCCGATCTGATTCTTCTGAAGATACCGTTAATAAGCACTTATTCAGAGT-666 721
QY      719 AGATCAAGAATTGTTCAAATACACC 744
Db      722 AGATCAAGAATTGTTCAAATACACC 747

```

## RESULT 12

```

Bf508987 739 bp mRNA linear EST 06-DEC-2000
LOCUS     Bf508987/c
DEFINITION UI-H-B14-aot-b-06-0-UI.s1 NCI-CGAP_Skb8 Homo sapiens cDNA clone
IMAGE:3085787 3', mRNA sequence.
Bf508987
ACCESSION Bf508987.1 GI:11592285
VERSION    EST
KEYWORDS   human.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

Db      1  TTAGCTTGACACAGGTTTGGAGTATGCAAAACAGCTAATAATTTTGCMAAAAGGAAATAA 60
Oy      458 CTCCTCGACATCTAAAGATGAAGTTTCATCATCCAAAGATGAGGGCTACAGAAACGG 517
Db      61 CTCCTCGACATCTAAAGATGAAGTTTCTATCATCCAAAGATGAGGGCTACAGAAACGG 120
Oy      518 TGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCCTTCTTGCAGAAAACAGTCTCAG 577
Db      121 TGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCCTTCTTGCAGAAAACAGTCTCAG 177
Oy      578 TGTCCAACTCTCTACACTTGGACACTGTGAACTCTGAGACAAAGACCGATACAC 637
Db      178 TGTCCAACTCTCTACACTTGGACACTGTGAACTCTGAGACAAAGACCGATACAC 237
Oy      638 TCMAAAGACGCTGCTCATATGAATGGATGATCTGATCTCTGAAAGTACCGTTAATA 697
Db      238 TCMAAAGACGCTGCTCATATGAATGGATGATCTGATCTCTGAAAGTACCGTTAATA 697
Oy      698 GGCACCTTATGCACTGTGGAGATCAAGAAATGTTACAAATCACCCCTCAAGAAACGAG 757
Db      265 ----- 264
Oy      758 GGATGAATTCAGTTTGATCTGCAAAAAGCGCTGCTTGAATTTTCTGAGACGATGT 817
Db      265 -----GGCTCTCTGTGAATTTTCTGAGACGATGT 294
Oy      818 AACAAATCTGACATCATCAACCCAGTAATATGATTTGAACACCACCTGAGAACGGCTGC 877
Db      295 AACAAATCTGACATCATCAACCCAGTAATATGATTTGAACACCACCTGAGAACGGCTGC 354
Oy      878 AGCTGAGAGGATCCAGAAAAGTATCAGGGTATCTGTTTCAAACTTGATGAGAGCC 937
Db      355 AGCTGAGAGGATCCAGAAAAGTATCAGGGTATCTGTTTCAAACTTGATGAGAGCC 414
Oy      938 ATGTGGGCAATATCTCATGCGACGCTCATTAAGATGAGACAGCAGTTTATCTAC 997
Db      415 ATGTGGGCAATATCTCATGCGACGCTCATTAAGATGAGACAGCAGTTTATCTAC 474
Oy      998 TAAAGACGAATGAATGTAGAAAAGGCTGAATTTCTGTAATAAACCAACCGCTGGCT 1057
Db      475 TAAAGACGAATGAATGTAGAAAAGGCTGAATTTCTGTAATAAACCAACCGCTGGCT 534
Oy      1058 AGCAAGAGCCAACTAAACAGATGGGCTGGAAGTAAGAAACATGTAATGATAGCGGAC 1117
Db      535 AGCAAGAGCCAACTAAACAGATGGGCTGGAAGTAAGAAACATGTAATGATAGCGGAC 594
Oy      1118 TCCGAGACAGAAAAAAGTAGATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATG 1177
Db      595 TCCGAGACAGAAAAAAGTAGATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATG 654
Oy      1178 GAATTAAGAGAAAGCTGCATGCTCAGAGAAATCTAGAGATCTGAAGATGTTCTTGAT 1237
Db      655 GAATTAAGAGAAAGCTGCATGCTCAGAGAAATCTAGAGATCTGAAGATGTTCTTGAT 714
Oy      1238 AACACTAAATAGCAGCATTTCAAGAAAGTTAATGAGTGGTTTCCAGAAAGTGAATGTT 1297
Db      715 AACACTAAATAGCAGCATTTCAAGAAAGTTAATGAGTGGTTTCCAGAAAGTGAATGTT 774
Oy      1298 AGGTTTCATGATCAGATGATGGGAGTCTGAATCAATGCCAAAGTACGT-GATGAT 1356
Db      775 AGGTTTCATGATCAGATGATGGGAGTCTGAATCAATGCCAAAGTACGTGAT 834
Oy      1357 TGACAGTT-CTAAATGAGTGAATG--AATATTCGTGGTCTTGAAGAAATTAACCTT 1411
Db      835 TGACAGTTCTTAATGAGTGAATGATTTCTGGTCTTTCAGAGAAATTAACCTT 893

```

```

RESULT 14
LOCUS   BF791668
DEFINITION 60225182.F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4344317 5',
ACCESSION BF791668

```

```

VERSION BF791668.1 GI:12096722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: LMAG962 row: k column: 06
High quality sequence stop: 678.

FEATURES
source
Location/Qualifiers
1..900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4344317"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-Sport6; Site: 1;
NotI; Site: 2; SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 283 a 176 c 212 g 229 t
ORIGIN
Query Match 11.5%; Score 657; DB 12; Length 900;
Best Local Similarity 96.5%; Pred. No. 3.4e-144;
Matches 736; Conservative 0; Mismatches 20; Indels 7; Gaps 6;

Oy      3475 CTGTTAATACAGATTTCTCTCCATATCTGATTTAGATTAAGTAAAGCTTATGGAA 3534
Db      1 CTGTTAATACAGATTTCTCTCCATATCTGATTTAGATTAAGTAAAGCTTATGGAA 60
Oy      3535 GTAGTCATGCAATCTAGGTTTCTTCTGAGACACCTGATGACCTGTATGATATGTGAAA 3594
Db      61 GTAGTCATGCAATCTAGGTTTCTTCTGAGACACCTGATGACCTGTATGATATGTGAAA 120
Oy      3595 TAAAGGAAGATCTAGTTTCTGTAAGTGAATTAAGAAAGTTCTGCTTTTGA 3654
Db      121 TAAAGGAAGATCTAGTTTCTGTAAGTGAATTAAGAAAGTTCTGCTTTTGA 180
Oy      3655 AAAGGTCACAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTTACCCATACACATTTGG 3714
Db      181 AAAGGTCACAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTTACCCATACACATTTGG 240
Oy      3715 CTCAGGTTTACGGAAGAGGGCCCAAGAAATTAAGATCTCAGAGAACTTATCTAGTG 3774
Db      241 CTCAGGTTTACGGAAGAGGGCCCAAGAAATTAAGATCTCAGAGAACTTATCTAGTG 300
Oy      3775 AGGATGAAGAGCTTCCCTGCTTCCAACTGTTATTTGTAAGTAAACATTTACCTT 3834
Db      301 AGGATGAAGAGCTTCCCTGCTTCCAACTGTTATTTGTAAGTAAACATTTACCTT 359
Oy      3835 CTCAGTCTACTAGGCAATAGACCGCTTGTCTACCGAGTGTCTCTTAAGAACACAGAGAGA 3894
Db      360 CTCAGTCTACTAGGCAATAGACCGCTTGTCTACCGAGTGTCTCTTAAGAACACAGAGAGA 419
Oy      3895 ATTATTTATCATTGAAGATAGCTTAATTAAGTCTGAGTAACACAGAGTAAATTTGGCAAGG 3954
Db      420 ATTATTTATCATTGAAGATAGCTTAATTAAGTCTGAGTAACACAGAGTAAAT-TGGCAAGG 478

```

QY 3955 CATCTCAGGAACATCACCCTAGTGAAGAAACAATATGTTCTGCTAGCTGTGTTTCTTCTAC 4014  
 Db 479 CATCTCAGGAACATCACCCTAGTGAAGAAACAATATGTTCTGCTAGCTGTGTTTCTTCTAC 538  
 QY 4015 AGTGCATGATGATGAGACTGTGCTCAATATCAACACCCAGGATCCCTTCTGATG 4074  
 Db 539 AGTGCATGATGATGAGACTGTGCTCAATATCAACACCCAGGATCCCTTCTGATG 598  
 QY 4075 GTTCTTCCAAACAATAGGAGCATGCTGTAAGCCAGGAGTTGGTCTGAGTGAAG 4134  
 Db 599 GTTCTTCCAAACAATAGGAGCATGCTGTAAGCCAGGAGTTGGTCTGAGTGAAG 656  
 QY 4135 AATTGTTTCAGATGATGAAGAAAGAGAACGGCTTGGAAGAAATATCAACAAGAGC 4194  
 Db 657 AA-TGGTTTCAGATGATGAAGAAAGAGAACGGCTTGGAAGAAATATCAACAAGAGC 714  
 QY 4195 AAACCATGATTCAACTAGTGAAGCAGCATGTGGGTGGA 4237  
 Db 715 CAAGCATGG-TTCACACTAGTGAAGCAGCATGTGGGTGGA 756

RESULT 15  
 LOCUS BG257190 938 bp mRNA linear EST 13-FEB-2001  
 DEFINITION 602377909F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4508450 5',  
 mRNA sequence.  
 ACCESSION BG257190  
 VERSION BG257190.1 GI:12767019  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 938)  
 NIH-MGC http://mgi.ncl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LLM10387 row: b column: 03  
 High quality sequence stop: 657.  
 Location/Qualifiers  
 1. 938  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4508450"  
 /clone\_id="NIH\_MGC\_92"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

FEATURES  
 source  
 BASE COUNT 294 a 193 c 225 g 226 t  
 ORIGIN

Query Match 11.4%; Score 651.6; DB 12; Length 938;  
 Best Local Similarity 93.9%; Pred. No. 6.4e-143;  
 Matches 712; Conservative 0; Mismatches 39; Indels 7; Gaps 3;

QY 30 CAGGCTGTGGGTTTCTCAGATACTGGGCGCTGCGCTCAGAGAGCCCTCACCCCTGC 89  
 Db 10 CAGGCTGTGGGTTTCTCAGATACTGGGCGCTGCGCTCAGAGAGCCCTCACCCCTGC 69  
 QY 90 TCTGGTAAAGTTCAATTGGAAGAAAGAAATGATTATCTGCTTCCGCTTGAAGAA 149

Db 70 TCTGGTAAAGTTCAATTGGAAGAAAGAAATGATTATCTGCTTCCGCTTGAAGAA 129  
 QY 150 GTACAAATGCTAATATGCTATGCTACGAAATCTTAGAGTGTCCATCTGCTGAGTGG 209  
 Db 130 GTACAAATGCTAATATGCTATGCTACGAAATCTTAGAGTGTCCATCTGCTGAGTGG 189  
 QY 210 ATCAGAGACCTGCTCCACAAAGTGTACACATATTTTGAAGATTTTGGATCTGAAA 269  
 Db 190 ATCAGAGACCTGCTCCACAAAGTGTACACATATTTTGAAGATTTTGGATCTGAAA 249  
 QY 270 CTTCYCAACGAGAAAGAGGCTTTCACAGTGTCTTTATGTAAGATGATATACCAA 329  
 Db 250 CTTCYCAACGAGAAAGAGGCTTTCACAGTGTCTTTATGTAAGATGATATACCAA 309  
 QY 330 AGAGACCTACAGAAAGATGAGATTTAGTCACTGTGTGAAGAGCTATTAATATCAT 389  
 Db 310 AGAGACCTACAGAAAGATGAGATTTAGTCACTGTGTGAAGAGCTATTAATATCAT 369  
 QY 390 TGTGCTTTTCAGCTTGAACAGAGTTTGAAGATGCAAAACAGCTATATTTTGAAGAA 449  
 Db 370 TGTGCTTTTCAGCTTGAACAGAGTTTGAAGATGCAAAACAGCTATATTTTGAAGAA 429  
 QY 450 GAAATATCTCTCTGGAACATCTAAAGATGAACTTCTATCATCAAAAGTATGGCTAC 509  
 Db 430 GAAATATCTCTCTGGAACATCTAAAGATGAACTTCTATCATCAAAAGTATGGCTAC 489  
 QY 510 AGAACCCTGCGCAAAAGACTTCTACAGATGAAACCCGAAATCTCTCTGCGAGAAAC 569  
 Db 490 AGAACCCTGCGCAAAAGACTTCTACAGATGAAACCCGAAATCTCTCTGCGAGAAAC 546  
 QY 570 AGTCTCAGTGTCCACTCTCTACCTTGAACCTGTGAACCTGTGAGACAAACAGAGCG 629  
 Db 547 AGTCTCAGTGTCCACTCTCTACCTTGAACCTGTGAGACAAACAGAGCGAGCG 606  
 QY 630 ATACACCTCAAAAGAGCTGTCTACATTTGAATTTGGATCTGATCTTCTG-AAAGTAC 688  
 Db 607 ATACACCTCAAAAGAGCTGTCTACATTTGAATTTGGATCTGATCTTCTG-AAAGTAC 663  
 QY 689 CGTTAATTAAGCAATATTGCACTGTGGGAGATCAAGAAATTTTACAAATACCCCTCA 748  
 Db 664 CGTTAATTAAGCAATATTGCACTGTGGGAGATCAAGAAATTTTACAAATACCCCT 723  
 QY 749 AGAACCAGGATGAATTCAGTTGGATTTGCAAAAA 786  
 Db 724 CAGGACAGGATGAATTCAGTTGGATTTGCAAAAA 761

Search completed: June 27, 2003, 04:59:57  
 Job time : 7230.45 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 16:33:51 ; Search time 734.849 Seconds  
(without alignments)  
11536.634 Million cell updates/sec

Title: US-09-734-672-3  
Perfect score: 5711  
Sequence: 1 AGCTCGCTGAGACTTCTCTG.....TCCCCACAGCAGCACTACTGA 5711

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	5709.4	100.0	5711	9	US-09-734-672-3
2	5709.4	100.0	5711	9	US-09-982-828-5
3	5704.6	99.9	5711	9	US-09-734-672-5
4	5704.6	99.9	5711	9	US-09-982-828-3
5	5704.6	99.9	5711	9	US-10-022-819-1
6	5703	99.9	5711	9	US-09-734-672-1
7	5703	99.9	5711	9	US-09-982-828-1
8	364.6	6.4	499	9	US-09-911-904-127
9	175.4	3.1	424	10	US-09-864-761-4552
10	147	2.6	147	10	US-09-864-761-21299
11	121	2.1	121	9	US-09-818-875-653
12	121	2.1	121	9	US-09-818-875-654
13	121	2.1	121	9	US-09-818-875-657
14	121	2.1	121	9	US-09-818-875-658
15	121	2.1	121	9	US-09-818-875-661
16	121	2.1	121	9	US-09-818-875-662
17	121	2.1	121	9	US-09-818-875-665
18	121	2.1	121	9	US-09-818-875-666
19	121	2.1	121	9	US-09-818-875-669

20	121	2.1	121	9	US-09-818-875-670	Sequence 670, App
21	121	2.1	121	9 <td>US-09-818-875-673</td> <td>Sequence 673, App</td>	US-09-818-875-673	Sequence 673, App
22	121	2.1	121	9 <td>US-09-818-875-674</td> <td>Sequence 674, App</td>	US-09-818-875-674	Sequence 674, App
23	121	2.1	121	9 <td>US-09-818-875-677</td> <td>Sequence 677, App</td>	US-09-818-875-677	Sequence 677, App
24	121	2.1	121	9 <td>US-09-818-875-678</td> <td>Sequence 678, App</td>	US-09-818-875-678	Sequence 678, App
25	121	2.1	121	9 <td>US-09-818-875-681</td> <td>Sequence 681, App</td>	US-09-818-875-681	Sequence 681, App
26	121	2.1	121	9 <td>US-09-818-875-682</td> <td>Sequence 682, App</td>	US-09-818-875-682	Sequence 682, App
27	121	2.1	121	9 <td>US-09-818-875-685</td> <td>Sequence 685, App</td>	US-09-818-875-685	Sequence 685, App
28	121	2.1	121	9 <td>US-09-818-875-686</td> <td>Sequence 686, App</td>	US-09-818-875-686	Sequence 686, App
29	121	2.1	121	9 <td>US-09-818-875-689</td> <td>Sequence 689, App</td>	US-09-818-875-689	Sequence 689, App
30	121	2.1	121	9 <td>US-09-818-875-690</td> <td>Sequence 690, App</td>	US-09-818-875-690	Sequence 690, App
31	121	2.1	121	9 <td>US-09-818-875-693</td> <td>Sequence 693, App</td>	US-09-818-875-693	Sequence 693, App
32	121	2.1	121	9 <td>US-09-818-875-694</td> <td>Sequence 694, App</td>	US-09-818-875-694	Sequence 694, App
33	121	2.1	121	9 <td>US-09-818-875-697</td> <td>Sequence 697, App</td>	US-09-818-875-697	Sequence 697, App
34	121	2.1	121	9 <td>US-09-818-875-698</td> <td>Sequence 698, App</td>	US-09-818-875-698	Sequence 698, App
35	121	2.1	121	9 <td>US-09-818-875-701</td> <td>Sequence 701, App</td>	US-09-818-875-701	Sequence 701, App
36	121	2.1	121	9 <td>US-09-818-875-702</td> <td>Sequence 702, App</td>	US-09-818-875-702	Sequence 702, App
37	121	2.1	121	9 <td>US-09-818-875-705</td> <td>Sequence 705, App</td>	US-09-818-875-705	Sequence 705, App
38	121	2.1	121	9 <td>US-09-818-875-706</td> <td>Sequence 706, App</td>	US-09-818-875-706	Sequence 706, App
39	121	2.1	121	9 <td>US-09-818-875-709</td> <td>Sequence 709, App</td>	US-09-818-875-709	Sequence 709, App
40	121	2.1	121	9 <td>US-09-818-875-710</td> <td>Sequence 710, App</td>	US-09-818-875-710	Sequence 710, App
41	121	2.1	121	9 <td>US-09-818-875-713</td> <td>Sequence 713, App</td>	US-09-818-875-713	Sequence 713, App
42	121	2.1	121	9 <td>US-09-818-875-714</td> <td>Sequence 714, App</td>	US-09-818-875-714	Sequence 714, App
43	121	2.1	121	9 <td>US-09-818-875-717</td> <td>Sequence 717, App</td>	US-09-818-875-717	Sequence 717, App
44	121	2.1	121	9 <td>US-09-818-875-718</td> <td>Sequence 718, App</td>	US-09-818-875-718	Sequence 718, App
45	121	2.1	121	9 <td>US-09-818-875-721</td> <td>Sequence 721, App</td>	US-09-818-875-721	Sequence 721, App

#### ALIGNMENTS

#### RESULT 1

US-09-734-672-3  
Sequence 3, Application US/09734672  
Publication No. US20020183268A1

#### GENERAL INFORMATION:

APPLICANT: MURPHY, Patricia D.  
Alvares, Antonette C.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin

#### TITLE OF INVENTION: Coding Sequences of the Human

#### BRCA1 Gene

NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington

STATE: District of Columbia  
COUNTRY: USA

ZIP: 20004

#### COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09734, 672  
FILING DATE: 03-Dec-2000

#### PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/966, 436  
FILING DATE: 07 No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598, 591  
FILING DATE: 12-Feb-96

#### ATTORNEY/AGENT INFORMATION:

NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCA1  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-734-672-3

Query Match 100.0%; Score 5709.4; DB 9; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCTCGCTGAGACTTCCTCGACCCCGCAGCCGCTGTGGGTTTCTCAGATACTGGGCC 60  
DB 1 AGCTCGCTGAGACTTCCTCGACCCCGCAGCCGCTGTGGGTTTCTCAGATACTGGGCC 60  
QY 61 CCTGCGCTCAGAGAGGCTTCCCTCTGCTGTGGTAAAGTCAATTGGAACAGAAAGAA 120  
DB 61 CCTGCGCTCAGAGAGGCTTCCCTCTGCTGTGGTAAAGTCAATTGGAACAGAAAGAA 120  
QY 121 TGGATTATCTGCTCTCGGCTGGAAGTAACAAATGTCATTAATGCTATGAGAAA 180  
DB 121 TGGATTATCTGCTCTCGGCTGGAAGTAACAAATGTCATTAATGCTATGAGAAA 180  
QY 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
DB 181 TCTTAGAGTGTCCCATCTGTCTGAGATTGATCAAGAACTGTCTCCCAAAAGTGTACC 240  
QY 241 ACATATTTTGCATAATTTTGGATCTGTAACCTTCTCAACAGAAAGGAGGCTTCAAGT 300  
DB 241 ACATATTTTGCATAATTTTGGATCTGTAACCTTCTCAACAGAAAGGAGGCTTCAAGT 300  
QY 301 GTCCCTTATGATAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
DB 301 GTCCCTTATGATAGATGATATACCAAAAGAGAGCTTCAAGAAAGTACGATTTAGTC 360  
QY 361 AACCTTGTGAAGAGCTATTGAAATCATTTTGTCTTTTCACTTGTGACACAGGTTTGGAGT 420  
DB 361 AACCTTGTGAAGAGCTATTGAAATCATTTTGTCTTTTCACTTGTGACACAGGTTTGGAGT 420  
QY 421 ATGCAAAAGCTATATTTTGCAGAAAGAAATTAACCTTCTGTAACATCTTAAGAGATG 480  
DB 421 ATGCAAAAGCTATATTTTGCAGAAAGAAATTAACCTTCTGTAACATCTTAAGAGATG 480  
QY 481 AAGTTTCTATCATCCAAAGATGAGGCTACAGAAACCGTGCAGAAAGCTTCAAGAGTG 540  
DB 481 AAGTTTCTATCATCCAAAGATGAGGCTACAGAAACCGTGCAGAAAGCTTCAAGAGTG 540  
QY 541 AACCCGAAAATCCCTTCTGTCAGAAACAGCTCTCAAGTCTCAACCTCTTAACCTTGGAA 600  
DB 541 AACCCGAAAATCCCTTCTGTCAGAAACAGCTCTCAAGTCTCAACCTCTTAACCTTGGAA 600  
QY 601 CTGTGGAAGCTCTGAGAGACAAAGAGGATACAACTCTCAAAAGAGCTGTCTACTATG 660  
DB 601 CTGTGGAAGCTCTGAGAGACAAAGAGGATACAACTCTCAAAAGAGCTGTCTACTATG 660  
QY 661 AATTGGATCTGATTTCTTCTGAGATACGCTTAATAAGGCACTTAATGAGTGGAG 720  
DB 661 AATTGGATCTGATTTCTTCTGAGATACGCTTAATAAGGCACTTAATGAGTGGAG 720  
QY 721 ATCAAGAAATTTTACAAATCACCCCTCAAGGAACAGAGATGAATCAAGTTGGATCTG 780  
DB 721 ATCAAGAAATTTTACAAATCACCCCTCAAGGAACAGAGATGAATCAAGTTGGATCTG 780  
QY 781 CAAAAAAGGCTGCTTGTGAATTTTCTGAGACGAGTGTAAACAAATCTGAACATCATCAAC 840

DB 781 CAAAAAAGGCTGCTTGTGAATTTTCTGAGACGAGTGTAAACAAATCTGAACATCATCAAC 840  
QY 841 CCAATATATATGATTTGAAACACACACAGAGGCTCAGCTGAGAGGATCCAGAAAGT 900  
DB 841 CCAATATATATGATTTGAAACACACAGAGGCTCAGCTGAGAGGATCCAGAAAGT 900  
QY 901 ATCAGGATGATCTGTTTCAAACTGATGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 960  
DB 901 ATCAGGATGATCTGTTTCAAACTGATGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 960  
QY 961 GCTCATTTACAGCATGAGAACAGCAGATTATCTCACTAAAGACAGAAATGATGAGAA 1020  
DB 961 GCTCATTTACAGCATGAGAACAGCAGATTATCTCACTAAAGACAGAAATGATGAGAA 1020  
QY 1021 AGGCTGATTTCTGATTAATAAGCAACAGCTGCTTACGAGGAGCCACATCAACAT 1080  
DB 1021 AGGCTGATTTCTGATTAATAAGCAACAGCTGCTTACGAGGAGCCACATCAACAT 1080  
QY 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGACTCCAGACAGAAAAAGGTAG 1140  
DB 1081 GGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGACTCCAGACAGAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAAATGAGCAAGAACTGCCATGCT 1200  
DB 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAAATGAGCAAGAACTGCCATGCT 1200  
QY 1201 CAGAGAACTCTAGAGATGATGAGATGCTGCTGATTAACACTTAATACCATTTTCA 1260  
DB 1201 CAGAGAACTCTAGAGATGATGAGATGCTGCTGATTAACACTTAATACCATTTTCA 1260  
QY 1261 AAGTTAATGATGATGTTTTCAGAAAGTGAATGATGTTAGTCTGATGATGATGATG 1320  
DB 1261 AAGTTAATGATGATGTTTTCAGAAAGTGAATGATGTTAGTCTGATGATGATGATG 1320  
QY 1321 GGGAGTCTGATTAATGCTCAAGATGCTGATGATGATGATGATGATGATGATGATG 1380  
DB 1321 GGGAGTCTGATTAATGCTCAAGATGCTGATGATGATGATGATGATGATGATGATG 1380  
QY 1381 AATATCTGCTTCTTCAAGAAATATGATGATGATGATGATGATGATGATGATGATG 1440  
DB 1381 AATATCTGCTTCTTCAAGAAATATGATGATGATGATGATGATGATGATGATGATG 1440  
QY 1441 TATGTAAAGTGAAGAGTTCACCTCAATCACTAGAGATGATGATGATGATGATGAT 1500  
DB 1441 TATGTAAAGTGAAGAGTTCACCTCAATCACTAGAGATGATGATGATGATGATGAT 1500  
QY 1501 TTGGGAAAGCTATTCGGAAGAAAGCAAGGCTCCCAACTTAAGCCATGTAACCTGA 1560  
DB 1501 TTGGGAAAGCTATTCGGAAGAAAGCAAGGCTCCCAACTTAAGCCATGTAACCTGA 1560  
QY 1561 TAATTTATGAGAGATTTGTTTACTGAGCCACAGATTAATACAGAGGCTCCCTCACA 1620  
DB 1561 TAATTTATGAGAGATTTGTTTACTGAGCCACAGATTAATACAGAGGCTCCCTCACA 1620  
QY 1621 AATTTAAGCGTAAAGAGAGACCTACATCAGGCTTCATCTGAGATTTTATCAAGAAAG 1680  
DB 1621 AATTTAAGCGTAAAGAGAGACCTACATCAGGCTTCATCTGAGATTTTATCAAGAAAG 1680  
QY 1681 CAGATTTGCGAGTTCAAAAGAGCTCTGGAATATATATCAGGAACTAACCAAGCGAGC 1740  
DB 1681 CAGATTTGCGAGTTCAAAAGAGCTCTGGAATATATATCAGGAACTAACCAAGCGAGC 1740  
QY 1741 AGAATGCTCAAGGATGATATTTACTAATAGTGGTATGAGATGAATGAATGAATGA 1800  
DB 1741 AGAATGCTCAAGGATGATATTTACTAATAGTGGTATGAGATGAATGAATGAATGA 1800  
QY 1801 CTATTCAGATGAGAAATTCCTAACCAATAGATCACTCCAGAAAGATCTGCTTCA 1860  
DB 1801 CTATTCAGATGAGAAATTCCTAACCAATAGATCACTCCAGAAAGATCTGCTTCA 1860  
QY 1861 AAAAGAAAGCTGACTATTAAGCAGCAGTATTAAGCAATATGGAATCAATTAATATCC 1920

Db 1861 AAGGAAAGCTGAACCTATAAGCAGCAGTATAGCAATATGAACTCGAATTAATATATC 1920  
 Qy 1921 ACAATTCAAAAGACACCTAAAAAGATAGGCTGAGAGAGAGTCTTACAGCAGCATATTC 1980  
 Db 1921 ACAATTCAAAAGACACCTAAAAAGATAGGCTGAGAGAGAGTCTTACAGCAGCATATTC 1980  
 Qy 1981 ATGGGCTTGAACCTAGTAGTCAAGTAAATCTAAGCCCACTAATTTGATCTGAATTTGCAA 2040  
 Db 1981 ATGGGCTTGAACCTAGTAGTCAAGTAAATCTAAGCCCACTAATTTGATCTGAATTTGCAA 2040  
 Qy 2041 TTGATAGTGTCTTACAGCAGTGAAGATTAAGAAAAAAGTACACCAATTCGACGCA 2100  
 Db 2041 TTGATAGTGTCTTACAGCAGTGAAGATTAAGAAAAAAGTACACCAATTCGACGCA 2100  
 Qy 2101 GGCACAGAGAACTCACTCACTGAGAGTAAAGAACTGCAACCTGAGACCAAGAA 2160  
 Db 2101 GGCACAGAGAACTCACTCACTGAGAGTAAAGAACTGCAACCTGAGACCAAGAA 2160  
 Qy 2161 GTAACAAAGCCAAATGAACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
 Db 2161 GTAACAAAGCCAAATGAACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTGA 2220  
 Qy 2221 AGTTAAACAATGACCTGCTGCTTCTTACTAAGTGTCAATATCCAGTGAATTTAAGAT 2280  
 Db 2221 AGTTAAACAATGACCTGCTGCTTCTTACTAAGTGTCAATATCCAGTGAATTTAAGAT 2280  
 Qy 2281 TTGTCATCTCAGGCTTCCAGAGAGAAAGAAAGAACTGAACCACTTAAAGTGT 2340  
 Db 2281 TTGTCATCTCAGGCTTCCAGAGAGAAAGAAAGAACTGAACCACTTAAAGTGT 2340  
 Qy 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCGCAAC 2400  
 Db 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCGCAAC 2400  
 Qy 2401 AAAATCTGTAGAGAGTGAAGCTGTAAGTCTGAGGAAAGGCAAAACGAAATCAAT 2460  
 Db 2401 AAAATCTGTAGAGAGTGAAGCTGTAAGTCTGAGGAAAGGCAAAACGAAATCAAT 2460  
 Qy 2461 AAATATCTGCTGTAAGTGAAGCTGTAAGTCTGAGGAAAGGCAAAACGAAATCAAT 2520  
 Db 2461 AAATATCTGCTGTAAGTGAAGCTGTAAGTCTGAGGAAAGGCAAAACGAAATCAAT 2520  
 Qy 2521 GTGTAGTCAAGTGTGACAGCATTTGAAAAACCCCAAGGCACTAATTCATGTTTCCAAAG 2580  
 Db 2521 GTGTAGTCAAGTGTGACAGCATTTGAAAAACCCCAAGGCACTAATTCATGTTTCCAAAG 2580  
 Qy 2581 ATATATGAATGACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGT 2640  
 Db 2581 ATATATGAATGACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGT 2640  
 Qy 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATGATGCTCAGTATTTGAGAAATCAT 2700  
 Db 2641 GGGAAACAGCATGGAATGGAAGAAAGTGAATGATGCTCAGTATTTGAGAAATCAT 2700  
 Qy 2701 TCAAGGTTTCAAAAGCCGACATGCTGCTGTTTCAAAATCCAGAAATGCAAGAGG 2760  
 Db 2701 TCAAGGTTTCAAAAGCCGACATGCTGCTGTTTCAAAATCCAGAAATGCAAGAGG 2760  
 Qy 2761 AATGTGCAACATTTCTGCCCCACTCTGGCTCTTAAAGAAAGTCCAAAGTCTCACT 2820  
 Db 2761 AATGTGCAACATTTCTGCCCCACTCTGGCTCTTAAAGAAAGTCCAAAGTCTCACT 2820  
 Qy 2821 TTGAATGTGAACAAAGAGAAATCAAGAAAGATGATCAATATCAAGCTGTAC 2880  
 Db 2821 TTGAATGTGAACAAAGAGAAATCAAGAAAGATGATCAATATCAAGCTGTAC 2880  
 Qy 2881 AGACAGTTAATATCATGAGAGGCTTCTGCTGTTGTCAGAAAGATGAAGCAATGTATA 2940  
 Db 2881 AGACAGTTAATATCATGAGAGGCTTCTGCTGTTGTCAGAAAGATGAAGCAATGTATA 2940  
 Qy 2941 ATGCAATATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTGAGAGCA 3000  
 Db 2941 ATGCAATATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTGAGAGCA 3000

Qy 3001 ACGAACTGACTCATTTACTCCAAATAAATGACCTTTTACAAAACCCATATGCTATAC 3060  
 Db 3001 ACGAACTGACTCATTTACTCCAAATAAATGACCTTTTACAAAACCCATATGCTATAC 3060  
 Qy 3061 CAGCATTCTTCCATCAAGTCAATTTGTTAAACTAAATGAAGAAAAATCTGCTAGAG 3120  
 Db 3061 CAGCATTCTTCCATCAAGTCAATTTGTTAAACTAAATGAAGAAAAATCTGCTAGAG 3120  
 Qy 3121 AAACTTTGAGAGAACATTCATGTCACCTGAAAGAAATGGGAATATGAAACATTTCCA 3180  
 Db 3121 AAACTTTGAGAGAACATTCATGTCACCTGAAAGAAATGGGAATATGAAACATTTCCA 3180  
 Qy 3181 GTACAGTGAACATTAAGCCTATATTAATTAAGAAATGTTTTTAAAGAGCAGCT 3240  
 Db 3181 GTACAGTGAACATTAAGCCTATATTAATTAAGAAATGTTTTTAAAGAGCAGCT 3240  
 Qy 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3300  
 Db 3241 CAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3300  
 Qy 3301 TAGGTTCCAGTGAATGAAGAAACATTCAGAGCAAGTGAAGAAAGAGGCAAAATTTGA 3360  
 Db 3301 TAGGTTCCAGTGAATGAAGAAACATTCAGAGCAAGTGAAGAAAGAGGCAAAATTTGA 3360  
 Qy 3361 ATGCTATGCTTAGATTAAGGAGGTTTTCAGACCTGAGGCTATTAACAAAGTCTCTGGA 3420  
 Db 3361 ATGCTATGCTTAGATTAAGGAGGTTTTCAGACCTGAGGCTATTAACAAAGTCTCTGGA 3420  
 Qy 3421 GTAATTTGTAAGCTCTGAATAAATAAAGCAAGAAATGAAGAGTACTGACACTGTTA 3480  
 Db 3421 GTAATTTGTAAGCTCTGAATAAATAAAGCAAGAAATGAAGAGTACTGACACTGTTA 3480  
 Qy 3481 ATACAGATTTCTCCATATCTGATTTACATTAAGAACACCTATGGAGTATGTC 3540  
 Db 3481 ATACAGATTTCTCCATATCTGATTTACATTAAGAACACCTATGGAGTATGTC 3540  
 Qy 3541 ATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTTGATGATGATGATGATGAT 3600  
 Db 3541 ATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTTGATGATGATGATGATGAT 3600  
 Qy 3601 AAGATACTAGTTTCTGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAAAGCG 3660  
 Db 3601 AAGATACTAGTTTCTGAAATGACATTAAGAAAGTCTGCTGTTTAAAGCAAAAGCG 3660  
 Qy 3661 TCCAGAAAGGAGACCTTAGAGAGGAGTCTTACCCCTTACACCATACATTTGGCTCAGG 3720  
 Db 3661 TCCAGAAAGGAGACCTTAGAGAGGAGTCTTACCCCTTACACCATACATTTGGCTCAGG 3720  
 Qy 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAACTATCTGAGAGATG 3780  
 Db 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAACTATCTGAGAGATG 3780  
 Qy 3781 AAGAGCTTCCCTGCTTCCAACTTTGTTTAAAGTAAACATTAATACCTCTCAGT 3840  
 Db 3781 AAGAGCTTCCCTGCTTCCAACTTTGTTTAAAGTAAACATTAATACCTCTCAGT 3840  
 Qy 3841 CTACTAGGCAATAGCAGGCTTCTACCGAGTGTCTGCTAGAACACAGAGAGAAATTTAT 3900  
 Db 3841 CTACTAGGCAATAGCAGGCTTCTACCGAGTGTCTGCTAGAACACAGAGAGAAATTTAT 3900  
 Qy 3901 TATCATTTGAAGATTAAGCTTTAATGACAGTAAACAGGATTAATTTGCAAGAGCATTC 3960  
 Db 3901 TATCATTTGAAGATTAAGCTTTAATGACAGTAAACAGGATTAATTTGCAAGAGCATTC 3960  
 Qy 3961 AGAACAATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTTTCTTACAGTGA 4020  
 Db 3961 AGAACAATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTTTCTTACAGTGA 4020  
 Qy 4021 GTGAAATTTGGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080  
 Db 4021 GTGAAATTTGGAAGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4080

QY	4081	CCAAACCAATGAGGCATCGTGTGAAAGGCACGAGGATGGTGTGAGTACAGAAATGG	4140
Db	4081	CCAAACCAATGAGGCATCGTGTGAAAGGCACGAGGATGGTGTGAGTACAGAAATGG	4140
QY	4141	TTTCAGATGATGAAGAAAGAGACGGGGCTTGGAAAGAAATATATCAAGAGACAAAGCA	4200
Db	4141	TTTCAGATGATGAAGAAAGAGACGGGGCTTGGAAAGAAATATATCAAGAGACAAAGCA	4200
QY	4201	TGGATTCAACTTAGTGAAGACGATCTGGGTGTGAGAGTGAACAAAGCGTCTTGAAG	4260
Db	4201	TGGATTCAACTTAGTGAAGACGATCTGGGTGTGAGAGTGAACAAAGCGTCTTGAAG	4260
QY	4261	ACTGCTCAGGGGTATCTCTCAGAGTGCATTTTAAACCACTGACAGAGGGATACCATGC	4320
Db	4261	ACTGCTCAGGGGTATCTCTCAGAGTGCATTTTAAACCACTGACAGAGGGATACCATGC	4320
QY	4321	AACATTAACCTGTATTAAGCTCCAGCAGAAATGGCTGAACATAGAACGTGTAGAACGC	4380
Db	4321	AACATTAACCTGTATTAAGCTCCAGCAGAAATGGCTGAACATAGAACGTGTAGAACGC	4380
QY	4381	ATGGAGGCCAGCGCTTCTAAGACGCTACCTTCCATATATAGTACGCTTCTGGCCCTTGAGG	4440
Db	4381	ATGGAGGCCAGCGCTTCTAAGACGCTACCTTCCATATATAGTACGCTTCTGGCCCTTGAGG	4440
QY	4441	ACCTCGCAAAATCCAGAACCAACACATCAGAAAAGACAGTATTAACCTTCACAGAAACTA	4500
Db	4441	ACCTCGCAAAATCCAGAACCAACACATCAGAAAAGACAGTATTAACCTTCACAGAAACTA	4500
QY	4501	GTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCCCTTCTGCTGACAAAGTTTGAGTGTG	4560
Db	4501	GTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCCCTTCTGCTGACAAAGTTTGAGTGTG	4560
QY	4561	CAGATAGTTCTACCGATAAAATTAAGAACCAGGAGTGGAAAGTCAATCCCTCTAAAT	4620
Db	4561	CAGATAGTTCTACCGATAAAATTAAGAACCAGGAGTGGAAAGTCAATCCCTCTAAAT	4620
QY	4621	GCCCATCATTTAGATATAGTGGTGTACATGACAGTATGCTGCTGGGAGTCTTCAGATTAAGA	4680
Db	4621	GCCCATCATTTAGATATAGTGGTGTACATGACAGTATGCTGCTGGGAGTCTTCAGATTAAGA	4680
QY	4681	ACTACCACATCTCAAGAGGAGCTCATTAAGTTTGTATGTGGAGGAGCAGACAGCTGGAG	4740
Db	4681	ACTACCACATCTCAAGAGGAGCTCATTAAGTTTGTATGTGGAGGAGCAGACAGCTGGAG	4740
QY	4741	AGTCTGGGCCCACACATTTGACGGAAACATCTTACTTGGCAAAGCAGACATCTAGAGGAA	4800
Db	4741	AGTCTGGGCCCACACATTTGACGGAAACATCTTACTTGGCAAAGCAGACATCTAGAGGAA	4800
QY	4801	CCCCTTACCTGGAATCTGGAATCAGCCCTCTCTGATGAGCCGTGAATCTGATCCTCTGTG	4860
Db	4801	CCCCTTACCTGGAATCTGGAATCAGCCCTCTCTGATGAGCCGTGAATCTGATCCTCTGTG	4860
QY	4861	AAGACAGAGCCCCAGAGTCAAGTCTGTTGGCAACATACATCTTCAACCTCTGCATTTGA	4920
Db	4861	AAGACAGAGCCCCAGAGTCAAGTCTGTTGGCAACATACATCTTCAACCTCTGCATTTGA	4920
QY	4921	AAAGTCCCAATTTGAAGTGTGCAGATCTGCCCAGAGTCCAGCTGCTCTCATACTACTG	4980
Db	4921	AAAGTCCCAATTTGAAGTGTGCAGATCTGCCCAGAGTCCAGCTGCTCTCATACTACTG	4980
QY	4981	ATACTGCTGGGTATATATGCAATGGAAGAAAGTGTGAGCAGGAGAGACCAATTTGACAG	5040
Db	4981	ATACTGCTGGGTATATATGCAATGGAAGAAAGTGTGAGCAGGAGAGACCAATTTGACAG	5040
QY	5041	CTTCAACAGAAAAGGCTCAACAAAGAAATGTCCATGTGGTGTGGCCCTGAGCCCAAGAG	5100
Db	5041	CTTCAACAGAAAAGGCTCAACAAAGAAATGTCCATGTGGTGTGGCCCTGAGCCCAAGAG	5100
QY	5101	AATTTATCTGTGTACAAAGTTTGGCAGAAAACACACATCACTTAATCTAATCTAATTA	5160
Db	5101	AATTTATCTGTGTGTACAAAGTTTGGCAGAAAACACACATCACTTAATCTAATCTAATTA	5160
QY	5161	CTGAAGAGACTACTCATCTTGTGTATGAAGAACAGATGCTGAGCTTTGTGTGTACAGGAGAC	5220

Db	5161	CTGAAGACACTCATCTGTTGTTATGAAGAACAGATGCTGACTTTGTGTGACGGACAC	5220
QY	5221	TGAATATTTTCTAGGAATTGGCGGAGAGAAATGGGTAGTACTATTTCTGGGTGACC	5280
Db	5221	TGAAATATTTTCTAGGAATGCGGAGAGAAATGGGTAGTACTATTTCTGGGTGACC	5280
QY	5281	AGTCTATTAAAGAAAGAAATGCTGATGTAGCATGATTTTGAATGACAGACATGTGG	5340
Db	5281	AGTCTATTAAAGAAAGAAATGCTGATGTAGCATGATTTTGAATGACAGACATGTGG	5340
QY	5341	TCAATGAGAAACACCAAGGTCCTAAGCAGACAGAGAAATCCAGAGACAGAAAGATCT	5400
Db	5341	TCAATGAGAAACACCAAGGTCCTAAGCAGACAGAGAAATCCAGAGACAGAAAGATCT	5400
QY	5401	TCAGGGGCTAGAAATCTGTGCTATGAGGCCCTTACCAACATGCCCCAGATCACTGG	5460
Db	5401	TCAGGGGCTAGAAATCTGTGCTATGAGGCCCTTACCAACATGCCCCAGATCACTGG	5460
QY	5461	AATGATGATGTCAGCTGTGTGCTGCTTCTGTGTGTAAGAGCTTTCATCATTCACCTTG	5520
Db	5461	AATGATGATGTCAGCTGTGTGCTGCTTCTGTGTGTAAGAGCTTTCATCATTCACCTTG	5520
QY	5521	GCACAGGTGTCACCACTTGTGTGCTGTGACAGCCAGATGCTTGACAGAGACATGCT	5580
Db	5521	GCACAGGTGTCACCACTTGTGTGCTGTGACAGCCAGATGCTTGACAGAGACATGCT	5580
QY	5581	TCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTTGACCCGAGAGTGGGTGTTGACA	5640
Db	5581	TCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTTGACCCGAGAGTGGGTGTTGACA	5640
QY	5641	GTTGAGCACTTACCACTGTCAGAGAGCTGTGACACTTACTATATCCAGATCCCCCACA	5700
Db	5641	GTTGAGCACTTACCACTGTCAGAGAGCTGTGACACTTACTATATCCAGATCCCCCACA	5700
QY	5701	GCCACTACTGA 5711	
Db	5701	GCCACTACTGA 5711	

FILING DATE: 1998-05-06  
 APPLICATION NUMBER: US 08/798,691  
 FILING DATE: 1997-02-12  
 APPLICATION NUMBER: US 08/598,591  
 FILING DATE: 1996-02-12  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michael S. Tuscan  
 REGISTRATION NUMBER: 43,210  
 REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-739-3000  
 TELEFAX: 202-739-3001  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5711 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: Linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 STRAIN: BRCAL (cm13)  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 17  
 MAP POSITION: 17q21  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-982-828-5

Query Match 100.0%; Score 5709.4; DB 9; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5710; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCTGGGTGAGTCTTCTGAGACCCGACCCAGCCTGAGGCTTTCACAGTACTGGGCC 60  
 DB 1 AGCTGGGTGAGTCTTCTGAGACCCGACCCAGCCTGAGGCTTTCACAGTACTGGGCC 60  
 QY 61 CCTGGCGTCAGAGGCGCTTCCACCTGCTGCTGGGTAAAGTTCATTGGAACAGAGAGAA 120  
 DB 61 CCTGGCGTCAGAGGCGCTTCCACCTGCTGCTGGGTAAAGTTCATTGGAACAGAGAGAA 120  
 QY 121 TGGATTATCTGCTCTTCGCGTGAAGAAATGATCAAAATGCTATATGCTATGACAGAAA 180  
 DB 121 TGGATTATCTGCTCTTCGCGTGAAGAAATGATCAAAATGCTATATGCTATGACAGAAA 180  
 QY 181 TCTTAGAGTGTCCATCTGCTGAGTGTATCAAGAACTGCTCCACAAAGTGTGACC 240  
 DB 181 TCTTAGAGTGTCCATCTGCTGAGTGTATCAAGAACTGCTCCACAAAGTGTGACC 240  
 QY 241 ACATATTTTGGCAATTTTGGCATGCTGAACTTCTCAACCGAGAGAAAGGCGCTTCACAGT 300  
 DB 241 ACATATTTTGGCAATTTTGGCATGCTGAACTTCTCAACCGAGAGAAAGGCGCTTCACAGT 300  
 QY 301 GTCTTTATGTAAAGATGATATACCAAAAGAGCCTACAGAAAGTACGAGATTTAGTC 360  
 DB 301 GTCTTTATGTAAAGATGATATACCAAAAGAGCCTACAGAAAGTACGAGATTTAGTC 360  
 QY 361 AACCTGTGAAGAGCTATTGAAATCATTTGCTTTTTCAGCTTGACACAGGTTGGAGT 420  
 DB 361 AACCTGTGAAGAGCTATTGAAATCATTTGCTTTTTCAGCTTGACACAGGTTGGAGT 420  
 QY 421 ATGCAAAAGCTATTAATTTTGGCAAAAGAGAAATTAATCTCTGTAACATCTTAAAGATG 480  
 DB 421 ATGCAAAAGCTATTAATTTTGGCAAAAGAGAAATTAATCTCTGTAACATCTTAAAGATG 480  
 QY 481 AAGTTTCTATCATCAAAAGATGAGCTACAGAAACCGTCCAAAGACCTTCAAGAGTG 540  
 DB 481 AAGTTTCTATCATCAAAAGATGAGCTACAGAAACCGTCCAAAGACCTTCAAGAGTG 540  
 QY 541 AACCCGAAATTCCTCTCTTCCAGAGAAACGAGTCTGAGTCCAACTCTCAACCTTGGAA 600  
 DB 541 AACCCGAAATTCCTCTCTTCCAGAGAAACGAGTCTGAGTCCAACTCTCAACCTTGGAA 600  
 QY 601 CTGTGAGAACTGTAGAGCAAAAGCAGCGGATATCAACCTCAAAAGAGCTGTCTACATTG 660

DB 601 CTGTGAGAACTGTAGAGCAAAAGCAGCGGATATCAACCTCAAAAGAGCTGTCTACATTG 660  
 QY 661 AATTGGAGTGTGATTTCTTCCAGAGATTAACCTTAATAAGCACTTATTTGAGTGGGAG 720  
 DB 661 AATTGGAGTGTGATTTCTTCCAGAGATTAACCTTAATAAGCACTTATTTGAGTGGGAG 720  
 QY 721 ATCAAGATTTGTAACAATCAACCCCAAGAGCCAGAGATGAATATGATGATTTCTG 780  
 DB 721 ATCAAGATTTGTAACAATCAACCCCAAGAGCCAGAGATGAATATGATGATTTCTG 780  
 QY 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGACGAGATGTAACAAATTAATCAATCAATCAAC 840  
 DB 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGACGAGATGTAACAAATTAATCAATCAATCAAC 840  
 QY 841 CCAGTATATATGATTTGTAACACCACTGAGAGGCTGAGTGAAGGCAATCAAGAAAGT 900  
 DB 841 CCAGTATATATGATTTGTAACACCACTGAGAGGCTGAGTGAAGGCAATCAAGAAAGT 900  
 QY 901 ATCAGGATGATTTGTTTCAACTGATGAGGCAATGAGCAGCAATATCAATCAATCAAC 960  
 DB 901 ATCAGGATGATTTGTTTCAACTGATGAGGCAATGAGCAGCAATATCAATCAATCAAC 960  
 QY 961 GCTCATTTACGATGAGAGACAGAGATTATTAATCACTAAGACAGAGATGATGAGAA 1020  
 DB 961 GCTCATTTACGATGAGAGACAGAGATTATTAATCACTAAGACAGAGATGATGAGAA 1020  
 QY 1021 AGGCTGATTTCTGATTAATAAGCAAGCAGCTGCTTACAGAGGAGCAGCAATCAACAGAT 1080  
 DB 1021 AGGCTGATTTCTGATTAATAAGCAAGCAGCTGCTTACAGAGGAGCAGCAATCAACAGAT 1080  
 QY 1081 GGGCTGGAATGAGAAACATGTAATGATGAGGAGATCCACAGACAGAAAGAGTAG 1140  
 DB 1081 GGGCTGGAATGAGAAACATGTAATGATGAGGAGATCCACAGACAGAAAGAGTAG 1140  
 QY 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGGAATTAAGCAGAACTGCAATGCT 1200  
 DB 1141 ATCTGAATGCTGATCCCTGCTGTGAGAGAAAGATGGAATTAAGCAGAACTGCAATGCT 1200  
 QY 1201 CAGAGAACTCTAAGAGATGATGAGATGCTCTTGGATACACTTAATAGCAGATTTAGA 1260  
 DB 1201 CAGAGAACTCTAAGAGATGATGAGATGCTCTTGGATACACTTAATAGCAGATTTAGA 1260  
 QY 1261 AAGTTATGATGATTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
 DB 1261 AAGTTATGATGATTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATG 1320  
 QY 1321 GGGAGTCTGATCAAAATGCCAAAGTACGATGATGATGATGATGATGATGATGATGATG 1380  
 DB 1321 GGGAGTCTGATCAAAATGCCAAAGTACGATGATGATGATGATGATGATGATGATGATG 1380  
 QY 1381 AATATTCTGCTTCTTCCAGAGAAATAGACTTACGAGATGATGATGATGATGATGATGATG 1440  
 DB 1381 AATATTCTGCTTCTTCCAGAGAAATAGACTTACGAGATGATGATGATGATGATGATGATG 1440  
 QY 1441 TATGTAAAGTGAAGAGTCACTCCAAATCAGTAGAGATTAATATGAGAGCAAAATAT 1500  
 DB 1441 TATGTAAAGTGAAGAGTCACTCCAAATCAGTAGAGATTAATATGAGAGCAAAATAT 1500  
 QY 1501 TTGGGAAACCTATCGGAGAGAGCAAGCCTCCCAATTAAGCATGATGATGATGATGATG 1560  
 DB 1501 TTGGGAAACCTATCGGAGAGAGCAAGCCTCCCAATTAAGCATGATGATGATGATGATG 1560  
 QY 1561 TAATTTATGAGAGATTTGTTTCTGAGAGCAGAGATTAATACAGAGCGTCCCTCAAAATA 1620  
 DB 1561 TAATTTATGAGAGATTTGTTTCTGAGAGCAGAGATTAATACAGAGCGTCCCTCAAAATA 1620  
 QY 1621 AATTTAAAGCTTAAGAGAGAGCTACATCAGAGCCTTCACTCTGAGAGATTTTATACAGAG 1680  
 DB 1621 AATTTAAAGCTTAAGAGAGAGCTACATCAGAGCCTTCACTCTGAGAGATTTTATACAGAG 1680  
 QY 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAAGGAGACTTAACCAAGCAGAGC 1740  
 DB 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAAGGAGACTTAACCAAGCAGAGC 1740

Dp	1681	CAGATTTGGCAGTTCCAAAAGCCTCCTGAAATGATTAATCAGGAAACTACCAAAAGCAGC	1740
Qy	1741	AGATGTGTCAGTGTGAATATATTACTAATTAAGTGTGTCATGACAATTAACAAAGGTGATT	1800
Dp	1741	AGAAATGGTCAAGTGTGAATATTACTAATTAAGTGTGTCATGACAATTAACAAAGGTGATT	1800
Qy	1801	CTATTCAAAATGAGAAAAATCTTAACCCCAATGAAATCACTCGTGAAGAAATCTGCTTTCA	1860
Dp	1801	CTATTCAAAATGAGAAAAATCTTAACCCCAATGAAATCACTCGTGAAGAAATCTGCTTTCA	1860
Qy	1861	AAAGCAAGCTGAACCTATTAAGCAGCAGTATTAGCAATATGAACTCGAATTAATATCC	1920
Dp	1861	AAACCAAAAGCTGAACCTATTAAGCAGCAGTATTAGCAATATGAACTCGAATTAATATCC	1920
Qy	1921	ACAATTTCAAAAGCACCCTAATAAAGAAATAGCGTAGAGGAGAACTCTTCTACCGGATATTC	1980
Dp	1921	ACAATTTCAAAAGCACCCTAATAAAGAAATAGCGTAGAGGAGAACTCTTCTACCGGATATTC	1980
Qy	1981	ATGCGCTTGAACCTAGTACTAGTAGAGAAATCTTAAGCCCACTTAATTTGTAAGTAAATGGAAA	2040
Dp	1981	ATGCGCTTGAACCTAGTACTAGTAGAGAAATCTTAAGCCCACTTAATTTGTAAGTAAATGGAAA	2040
Qy	2041	TTGATAGTGTGTTCTTAGCAGTGAAGAGATTAAGAAAAAAAGTAAACCAAAATGGCAGTCA	2100
Dp	2041	TTGATAGTGTGTTCTTAGCAGTGAAGAGATTAAGAAAAAAAGTAAACCAAAATGGCAGTCA	2100
Qy	2101	GGCAGCAGCAAAACCTTACAACTCAATGGAAGTAAAGAACTGCAACTGAGAGCCAAAGAGA	2160
Dp	2101	GGCAGCAGCAAAACCTTACAACTCAATGGAAGTAAAGAACTGCAACTGAGAGCCAAAGAGA	2160
Qy	2161	GTAACAGAGCCAAATGAGACAGACAAAGTAAGAATGACATGATGATATCTTTCCAGAGCTGA	2220
Dp	2161	GTAACAGAGCCAAATGAGACAGACAAAGTAAGAATGACATGATGATATCTTTCCAGAGCTGA	2220
Qy	2221	AGTTAACAAATGACACCTGGTCTTCTTCTTAAGTGTTCAAATACAGAGAACTTAAAGAT	2280
Dp	2221	AGTTAACAAATGACACCTGGTCTTCTTCTTACTAAGTGTTCAAATACAGAGAACTTAAAGAT	2280
Qy	2281	TTGTCAATCTTAGCCTTCCAAAGAGAAAAAGAAAGTAAGAACTGAAACAGTTAAAGTGT	2340
Dp	2281	TTGTCAATCTTAGCCTTCCAAAGAGAAAAAGAAAGTAAGAACTGAAACAGTTAAAGTGT	2340
Qy	2341	CTAATTAATGCTGAAGAGCCCCAAAGATGTCATGTTAAGGGGAAGGGTTTGGCAAACTG	2400
Dp	2341	CTAATTAATGCTGAAGAGCCCCAAAGATGTCATGTTAAGGGGAAGGGTTTGGCAAACTG	2400
Qy	2401	AAAGATCTGTAGAGAGTAGACAGATATTCATTTGTAAGTGTGTAAGTATTAAGCACTCAG	2460
Dp	2401	AAAGATCTGTAGAGAGTAGACAGATATTCATTTGTAAGTGTGTAAGTATTAAGCACTCAG	2460
Qy	2461	AAAGATCTGTAGAGAGTAGACAGATATTCATTTGTAAGTGTGTAAGTATTAAGCACTCAG	2520
Dp	2461	AAAGATCTGTAGAGAGTAGACAGATATTCATTTGTAAGTGTGTAAGTATTAAGCACTCAG	2520
Qy	2521	GTGTGAGTCACTGTGACAGCATTTTGAATAACCCCAAGGAGCTAATTCATGGTTGTTCCAAAG	2580
Dp	2521	GTGTGAGTCACTGTGACAGCATTTTGAATAACCCCAAGGAGCTAATTCATGGTTGTTCCAAAG	2580
Qy	2581	ATAATAGAAATGACACAGAGAGCTTTAAGTATTCATTGGGACATGAAGTTAACCACAGTC	2640
Dp	2581	ATAATAGAAATGACACAGAGAGCTTTAAGTATTCATTGGGACATGAAGTTAACCACAGTC	2640
Qy	2641	GGGAAACAGAGATGGAATGGAAAGTAAGTAAGTGAATGTCAGTATTTTCCAGATATCAT	2700
Dp	2641	GGGAAACAGAGATGGAATGGAAAGTAAGTAAGTGAATGTCAGTATTTTCCAGATATCAT	2700
Qy	2701	TCAAGGTTTCAAAAGCGCCAGTCAATTTGCTGTGTTTAAATCCAGAAATGCAAGAGAG	2760
Dp	2701	TCAAGGTTTCAAAAGCGCCAGTCAATTTGCTGTGTTTAAATCCAGAAATGCAAGAGAG	2760
Qy	2761	AAATGTGCAACATTTCTGTGGCCACTCTGTGGGTCTTTAAAGAAACAAAGTCAACTT	2820
Dp	2761	AAATGTGCAACATTTCTGTGGCCACTCTGTGGGTCTTTAAAGAAACAAAGTCAACTT	2820

OY	2821	TTGAATGTGAACAAAGGAAGAAATCAAGGAAAGAAATGAGCTTAATTCAGGCTGTAC	2880
Db	2821	TTGATATGTGAACAAAGGAAGAAATCAAGGAAAGAAATGAGCTTAATTCAGGCTGTAC	2880
OY	2881	AGACAGTTAAATATCAGTCGAGGCTTCTGTGGTGGTCAGAAAGATAGCCAGCTGATA	2940
Db	2881	AGACAGTTAAATATCAGTCGAGGCTTCTGTGGTGGTCAGAAAGATAGCCAGCTGATA	2940
OY	2941	ATGCGAAATGACTATATCAAAAGGAGGCTGAGTTTTGTCTATCATCTCAGTTTCAGAGCA	3000
Db	2941	ATGCGAAATGACTATATCAAAAGGAGGCTGAGTTTTGTCTATCATCTCAGTTTCAGAGCA	3000
OY	3001	ACGAAACTGACATCTACTCTCCAAATAAACAATGACCTTTTCAAAACCATATGTTATAC	3060
Db	3001	ACGAAACTGACATCTACTCTCCAAATAAACAATGACCTTTTCAAAACCATATGTTATAC	3060
OY	3061	CACCACTTTTCCCATCAAGTCACTATTTGTTAAACTAAATGTATAGAAAATCTGCTAGAG	3120
Db	3061	CACCACTTTTCCCATCAAGTCACTATTTGTTAAACTAAATGTATAGAAAATCTGCTAGAG	3120
OY	3121	AAACCTTTGGAGAACCTTCATATGTCACTCGAAGAGAAATGGGAAATGAGAACCTTCCAA	3180
Db	3121	AAACCTTTGGAGAACCTTCATATGTCACTCGAAGAGAAATGGGAAATGAGAACCTTCCAA	3180
OY	3181	GTAACGTGAGCACATTTAGCCGTAATACATTAGAGAAATGTTTTTAAAGAACCCAGCT	3240
Db	3181	GTAACGTGAGCACATTTAGCCGTAATACATTAGAGAAATGTTTTTAAAGAACCCAGCT	3240
OY	3241	CAAGCAATTTAATGAAGTAGTGTCAGTCACTAAATGAAGTGGGCTCCAGATTAAATGAAA	3300
Db	3241	CAAGCAATTTAATGAAGTAGTGTCAGTCACTAAATGAAGTGGGCTCCAGATTAAATGAAA	3300
OY	3301	TAGGTTCCAGTGTGAACAACTTCAAGCAGAACTAGTAGAAGACAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTGTGAACAACTTCAAGCAGAACTAGTAGAAGACAGAGGGCCAAATTTGA	3360
OY	3361	ATGCTATNGCTTAATATGAGGGTTTTGCAACCTGAGCTCTATTAACAAAGTCTTCTGGAA	3420
Db	3361	ATGCTATNGCTTAATATGAGGGTTTTGCAACCTGAGCTCTATTAACAAAGTCTTCTGGAA	3420
OY	3421	GTAATTTGAAGCATCCTGAAATPAAAAAAGCAAGATATGAAGAGTGGTTCAGACTGTAA	3480
Db	3421	GTAATTTGAAGCATCCTGAAATPAAAAAAGCAAGATATGAAGAGTGGTTCAGACTGTAA	3480
OY	3481	ATACAGATTTCTCTCCATATCTGATTTCAAGTAACTTAAGAACGCCATATGGAAGTAGTC	3540
Db	3481	ATACAGATTTCTCTCCATATCTGATTTCAAGTAACTTAAGAACGCCATATGGAAGTAGTC	3540
OY	3541	ATACATCTCAGGTTTGTCTGAGAACCGATGACCTTGTATGATGATGGTGAATTAAGG	3600
Db	3541	ATGCACTCTCAGGTTTGTCTGAGAACCGTGAATGACCTGTTATGATGATGGTGAATTAAGG	3600
OY	3601	AAAGATCTAGTTTTGCTGAAATGACATTAAAGAAAGTTCTGCTGTTTTAGCAAAAAGCG	3660
Db	3601	AAAGATCTAGTTTTGCTGAAATGACATTAAAGAAAGTTCTGCTGTTTTAGCAAAAAGCG	3660
OY	3661	TCGAGAAAGAGAGGCTTACAGAGACTCTAGCCCTTACACCCATATACATTTGGCTAGG	3720
Db	3661	TCGAGAAAGAGAGGCTTACAGAGACTCTAGCCCTTACACCCATATACATTTGGCTAGG	3720
OY	3721	GTTATCCGAGAGAGGGGCAAGAAATTAAGTCTCTCAGAAAGAACATTATCTATGAGAGATG	3780
Db	3721	GTTATCCGAGAGAGGGGCAAGAAATTAAGTCTCTCAGAAAGAACATTATCTATGAGAGATG	3780
OY	3781	AAAGAGCTTCCCTGCTTCCACACCTTGTATTTGGTAAAGTAACAAATATACCTTCTCAGT	3840
Db	3781	AAAGAGCTTCCCTGCTTCCACACCTTGTATTTGGTAAAGTAACAAATATACCTTCTCAGT	3840
OY	3841	CTACTATGGCAATGACACCGTTGCTACCGAGGTGTCTGTCTAAACACAGAGAGAAATTTAT	3900
Db	3841	CTACTATGGCAATGACACCGTTGCTACCGAGGTGTCTGTCTAAACACAGAGAGAAATTTAT	3900

QY 3901 TATCATGAGATAGTCTTAATGACTGCTAGTACCAAGGTAATATGCGCAAGGCAATCTC 3960  
 Db 3901 TATCATGAGATAGTCTTAATGACTGCTAGTACCAAGGTAATATGCGCAAGGCAATCTC 3960  
 QY 3961 AGGAATCATCCTAGTGGAGAAACAAATGTTCTGCTAGTCTGTTCTTCTACAGTGA 4020  
 Db 3961 AGGAATCATCCTAGTGGAGAAACAAATGTTCTGCTAGTCTGTTCTTCTACAGTGA 4020  
 QY 4021 GTGATTTGGAAGACTTGTACTGCAAAATACAAACCCAGATCCCTTCTGATGTTCTT 4080  
 Db 4021 GTGATTTGGAAGACTTGTACTGCAAAATACAAACCCAGATCCCTTCTGATGTTCTT 4080  
 QY 4081 CCAATCAATGAGGCTAGTCTGTAAGGCAAGGAGTGTGCTGAGTACAGGAATGG 4140  
 Db 4081 CCAATCAATGAGGCTAGTCTGTAAGGCAAGGAGTGTGCTGAGTACAGGAATGG 4140  
 QY 4141 TTTGATGATGAGGAAGGAAGGAGGCTTGGAGGAATAATACAGAAAGGAAGCA 4200  
 Db 4141 TTTGATGATGAGGAAGGAAGGAGGCTTGGAGGAATAATACAGAAAGGAAGCA 4200  
 QY 4201 TGGATTTCAACTTGTAGTGAAGCAGATCTGGGTGTGAGAGTGAAGCAAGCTCTGAG 4260  
 Db 4201 TGGATTTCAACTTGTAGTGAAGCAGATCTGGGTGTGAGAGTGAAGCAAGCTCTGAG 4260  
 QY 4261 ACTGCTCAGGCTATCTCTCAGAGTACATTTTAACTTACCTCAGCAGAGGATACATGC 4320  
 Db 4261 ACTGCTCAGGCTATCTCTCAGAGTACATTTTAACTTACCTCAGCAGAGGATACATGC 4320  
 QY 4321 AACATACCTGATTAACCTCCAGAGAAATGCTGAACATTAAGAGCTGTGTAAACAGC 4380  
 Db 4321 AACATACCTGATTAACCTCCAGAGAAATGCTGAACATTAAGAGCTGTGTAAACAGC 4380  
 QY 4381 ATGGAGCAGCAGCTTCTACAGTACCTTCCATCATTAAGTACCTTCTGCTGAGG 4440  
 Db 4381 ATGGAGCAGCAGCTTCTACAGTACCTTCCATCATTAAGTACCTTCTGCTGAGG 4440  
 QY 4441 ACCTGCAAAATCCAGAAACAAAGCAGATCAGAAAAAGCAGTAATTAACCTCAGCAAGGA 4500  
 Db 4441 ACCTGCAAAATCCAGAAACAAAGCAGATCAGAAAAAGCAGTAATTAACCTCAGCAAGGA 4500  
 QY 4501 GTGATTAACCTTATAGGCAAGATCCAGAAAGCCTTCTGCTGACAACTTTGAGGTGCTG 4560  
 Db 4501 GTGATTAACCTTATAGGCAAGATCCAGAAAGCCTTCTGCTGACAACTTTGAGGTGCTG 4560  
 QY 4561 CAGATAGTCTTACAGTAAATTAAGAACAGAGAGTGAAGAAAGTCACTCCCTCTTAAT 4620  
 Db 4561 CAGATAGTCTTACAGTAAATTAAGAACAGAGAGTGAAGAAAGTCACTCCCTCTTAAT 4620  
 QY 4621 GCCCATATTAGATGATAGGTGTATGACAGATGCTGAGAGTCTTCAAGATGAA 4680  
 Db 4621 GCCCATATTAGATGATAGGTGTATGACAGATGCTGAGAGTCTTCAAGATGAA 4680  
 QY 4681 ACTTACCCTTCAAGAGAGCTCTTAAAGTGTGTATGTGAGAGAGCAACAGCTGGAAG 4740  
 Db 4681 ACTTACCCTTCAAGAGAGCTCTTAAAGTGTGTATGTGAGAGAGCAACAGCTGGAAG 4740  
 QY 4741 AGTGTGGGCAACAGATTTGAGGAAACATCTTACTTGGCAAGCAAGATCTAGAGGAA 4800  
 Db 4741 AGTGTGGGCAACAGATTTGAGGAAACATCTTACTTGGCAAGCAAGATCTAGAGGAA 4800  
 QY 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGACCTGAACTGATCTTCTG 4860  
 Db 4801 CCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGACCTGAACTGATCTTCTG 4860  
 QY 4861 AAGACAGAGCCAGATCTGAGTCTGTGTGGCAACATACCATCTTCAACCTCTGCAATGA 4920  
 Db 4861 AAGACAGAGCCAGATCTGAGTCTGTGTGGCAACATACCATCTTCAACCTCTGCAATGA 4920  
 QY 4921 AAGTCCCAATGAAAGTGTGAGAAATCTCCAGAGTCTGAGTCTGATCTACTG 4980  
 Db 4921 AAGTCCCAATGAAAGTGTGAGAAATCTCCAGAGTCTGAGTCTGATCTACTG 4980  
 QY 4981 ATACTGCTGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGGAAGCCAGAAATGACAG 5040

Db 4981 ATACTGCTGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGGAAGCCAGAAATGACAG 5040  
 QY 5041 CTTCACAGAAAGGCTCAACAAAGAAATGCTCATGGTGTGCTGAGCCTGAGCCAGAG 5100  
 Db 5041 CTTCACAGAAAGGCTCAACAAAGAAATGCTCATGGTGTGCTGAGCCTGAGCCAGAG 5100  
 QY 5101 AATTATGCTGCTGATACAGTTTGGCAGAAACACCATCATCTTAACTTAATTA 5160  
 Db 5101 AATTATGCTGCTGATACAGTTTGGCAGAAACACCATCATCTTAACTTAATTA 5160  
 QY 5161 CTGAAGACACTACATGATGTTTATGAACAGATGCTGAGTTGTGTGACAGGAC 5220  
 Db 5161 CTGAAGACACTACATGATGTTTATGAACAGATGCTGAGTTGTGTGACAGGAC 5220  
 QY 5221 TGAATATTTCTGGAATGCTGGAGGAGAAATGAGTGTGATTTCTGGGGGAC 5280  
 Db 5221 TGAATATTTCTGGAATGCTGGAGGAGAAATGAGTGTGATTTCTGGGGGAC 5280  
 QY 5281 AGCTATTTAAAGAAAGAAATGCTGATGAGCATGATTTGAAGTCAAGAGATG 5340  
 Db 5281 AGCTATTTAAAGAAAGAAATGCTGATGAGCATGATTTGAAGTCAAGAGATG 5340  
 QY 5341 TCAATGAGAAACCCAGAGTCCAAAGCAGAGAGAAATCCAGAGCAGAAAGATCT 5400  
 Db 5341 TCAATGAGAAACCCAGAGTCCAAAGCAGAGAGAAATCCAGAGCAGAAAGATCT 5400  
 QY 5401 TCAGGGGCTGGAATGCTGATGAGGAGGCTTCCACCAATGCTCCACAGATCACTG 5460  
 Db 5401 TCAGGGGCTGGAATGCTGATGAGGAGGCTTCCACCAATGCTCCACAGATCACTG 5460  
 QY 5461 AATGATGATGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5520  
 Db 5461 AATGATGATGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5520  
 QY 5521 GCACAGGTGCTCACCCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5580  
 Db 5521 GCACAGGTGCTCACCCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5580  
 QY 5581 TCCATGCAATGGGCAAGTGTGAGGAGCAGCTGTGCTGCTGCTGCTGCTGCTG 5640  
 Db 5581 TCCATGCAATGGGCAAGTGTGAGGAGCAGCTGTGCTGCTGCTGCTGCTGCTG 5640  
 QY 5641 GTGTAGCACTTACAGAGTCCAGAGAGTGTGAGGAGCAGCTGTGCTGCTGCTG 5700  
 Db 5641 GTGTAGCACTTACAGAGTCCAGAGAGTGTGAGGAGCAGCTGTGCTGCTGCTG 5700  
 QY 5701 GCCACTACTGA 5711  
 Db 5701 GCCACTACTGA 5711

RESULT 3  
 US-09-734-672-5  
 : Sequence 5, Application US/09734672  
 : Publication No. US20020183268A1  
 GENERAL INFORMATION:  
 APPLICANT: Murphy, Patricia D.  
 Allen, Antonette C.  
 Alvares, Christopher P.  
 Critz, Brenda S.  
 Olson, Sheri J.  
 Schelter, Denise B.  
 Zeng, Bin  
 TITLE OF INVENTION: Coding Sequences of the Human  
 BRCA1 Gene  
 NUMBER OF SEQUENCES: 72  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morgan Lewis & Bockius LLP  
 STREET: 1111 Pennsylvania Ave., N.W.  
 CITY: Washington  
 STATE: District of Columbia  
 COUNTRY: USA







D	1561	TAATATTAAGGAGCATTTGTTACTAGGCCACAGATAATACAAAGACGTCCCTCCACAATA	1620
O	1621	AATTAAGCCCTAAAGAGACCTACATCAGGCCCTTCATCCCTGAGGATTTTTCAGAAAG	1680
D	1621	AATTAAGCCCTAAAGAGACCTACATCAGGCCCTTCATCCCTGAGGATTTTTCAGAAAG	1680
O	1681	CAGATTTGGCAGTTCAAAGAGACCTCGTGAATGATTAATCAGGAACTAACCAACGAGC	1740
D	1681	CAGATTTGGCAGTTCAAAGAGACCTCGTGAATGATTAATCAGGAACTAACCAACGAGC	1740
O	1741	AGAAATGCTCAAGTATGATTAATTAATTAATGCTCATGAGAAATTAACAAAGGTGATT	1800
D	1741	AGAAATGCTCAAGTATGATTAATTAATTAATGCTCATGAGAAATTAACAAAGGTGATT	1800
O	1801	CTATTCGAATGAGAAAAATCTCAACCAATAGATCAGTCAAAAGAAATCTGCTTCA	1860
D	1801	CTATTCGAATGAGAAAAATCTCAACCAATAGATCAGTCAAAAGAAATCTGCTTCA	1860
O	1861	AAACGAAAGCTGAACCTATTAAGCAGCAGTATAGCAATATGGAATCGAATTAATATCC	1920
D	1861	AAACGAAAGCTGAACCTATTAAGCAGCAGTATAGCAATATGGAATCGAATTAATATCC	1920
O	1921	ACAATTCAAAAGCAGCTTAAGAAAGATAGGCTGAGAGAGAGCTCTTACCAGCATATTC	1980
D	1921	ACAATTCAAAAGCAGCTTAAGAAAGATAGGCTGAGAGAGAGCTCTTACCAGCATATTC	1980
O	1981	ATGGGCTTGAATAGTATGCTAGTGAATACTCAACCCCACTAATTTGATGATTCGAAA	2040
D	1981	ATGGGCTTGAATAGTATGCTAGTGAATACTCAACCCCACTAATTTGATGATTCGAAA	2040
O	2041	TTGATAGTGTGTTAGCAGTGAAGATTAAGAAAAAAGTACACCAATGGCAGTCA	2100
D	2041	TTGATAGTGTGTTAGCAGTGAAGATTAAGAAAAAAGTACACCAATGGCAGTCA	2100
O	2101	GGCAGACGAGAACCTTACACTCATGTAAGAGTAAAGAACTGCACTGGAGCCAAAGAA	2160
D	2101	GGCAGACGAGAACCTTACACTCATGTAAGAGTAAAGAACTGCACTGGAGCCAAAGAA	2160
O	2161	GTACAAGCCAAATGAACAGACAGTAAGAAAGCATGACAGTACTTCCACAGCTGA	2220
D	2161	GTACAAGCCAAATGAACAGACAGTAAGAAAGCATGACAGTACTTCCACAGCTGA	2220
O	2221	AGTTAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATACAGTGAATTAAGAAAT	2280
D	2221	AGTTAACAAATGACCTGTTCTTTTACTAAGTGTCAAAATACAGTGAATTAAGAAAT	2280
O	2281	TTGTCATCTCCTAGCTTCCAGAGAGAGAAAAAGAAAGAACTGAAACAGTTAAAGTGT	2340
D	2281	TTGTCATCTCCTAGCTTCCAGAGAGAGAAAAAGAAAGAACTGAAACAGTTAAAGTGT	2340
O	2341	CTATATATGCTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGGTTTTCGAACTG	2400
D	2341	CTATATATGCTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGGTTTTCGAACTG	2400
O	2401	AAAGATCTGAGAGTAGCAGTATTTCAATGCTGACTGCTACTGATTAATGAGCAGCTCAGG	2460
D	2401	AAAGATCTGAGAGTAGCAGTATTTCAATGCTGACTGCTACTGATTAATGAGCAGCTCAGG	2460
O	2461	AAAGATCTGCTTACTGGAAGTTAGCACTGTAGGGAAGGCAAAAACAGAACTCAATTAAT	2520
D	2461	AAAGATCTGCTTACTGGAAGTTAGCACTGTAGGGAAGGCAAAAACAGAACTCAATTAAT	2520
O	2521	GTGAGATCAGTGGCGAGCATTTGAAAGCCCAAGGAGCTAATTCATGATTTTCGCAAG	2580
D	2521	GTGAGATCAGTGGCGAGCATTTGAAAGCCCAAGGAGCTAATTCATGATTTTCGCAAG	2580
O	2581	ATAATGAATGACACAGAGAGGCTTTAAGTATCCATTTGGAGCATGAAGTTAACACAGTC	2640
D	2581	ATAATGAATGACACAGAGAGGCTTTAAGTATCCATTTGGAGCATGAAGTTAACACAGTC	2640
O	2641	GGGAACACAGCATRGAATGGAAGAAAGTAACCTTGATGCTCAGTATTTTCAGAAATCAT	2700
D	2641	GGGAACACAGCATRGAATGGAAGAAAGTAACCTTGATGCTCAGTATTTTCAGAAATCAT	2700
D	2641	GGGAACACAGCATRGAATGGAAGAAAGTAACCTTGATGCTCAGTATTTTCAGAAATCAT	2700
O	2701	TCAGGTTTCAAGGGCCAGCATTTGCTGTTTTCGTTTCAATCCAGGAATGCAGAGAGG	2760
D	2701	TCAGGTTTCAAGGGCCAGCATTTGCTGTTTTCGTTTCAATCCAGGAATGCAGAGAGG	2760
O	2761	AATGTCACACATTTCTCCCTCTGCGGCTTTTAAAGAAACAAAGTCCAAAGCAGCTT	2820
D	2761	AATGTCACACATTTCTCCCTCTGCGGCTTTTAAAGAAACAAAGTCCAAAGCAGCTT	2820
O	2821	TTGAATGTGAACAAAGAGAGAAAAATCAAGAAAGATGACTAATATCAAGCCTGTAC	2880
D	2821	TTGAATGTGAACAAAGAGAGAAAAATCAAGAAAGATGACTAATATCAAGCCTGTAC	2880
O	2881	AGACAGTTAATATCAGTGCAGGCTTCTGCTGCTTGCACAAAAGATTAAGCAGTTGTA	2940
D	2881	AGACAGTTAATATCAGTGCAGGCTTCTGCTGCTTGCACAAAAGATTAAGCAGTTGTA	2940
O	2941	ATGCCAAATGTAGTATCAAGAGAGCTGTAGTGTTCATCATCTCAGTTTCAGAGGCA	3000
D	2941	ATGCCAAATGTAGTATCAAGAGAGCTGTAGTGTTCATCATCTCAGTTTCAGAGGCA	3000
O	3001	ACGAACCTGACATCTTCAATTAACATGACATTTTCAAAACCCATATGCTATAC	3060
D	3001	ACGAACCTGACATCTTCAATTAACATGACATTTTCAAAACCCATATGCTATAC	3060
O	3061	CACCACCTTTTCCATCAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAGG	3120
D	3061	CACCACCTTTTCCATCAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAGG	3120
O	3121	AAAACCTTGGAGACATTCATGTCACCTGGAAGAGAAATGGAAATGAGAACTTCCAA	3180
D	3121	AAAACCTTGGAGACATTCATGTCACCTGGAAGAGAAATGGAAATGAGAACTTCCAA	3180
O	3181	GTACAGTGAGCACAATTAGCCGTAAATTAAGAAAAATGTTTAAAGAGCAGACT	3240
D	3181	GTACAGTGAGCACAATTAGCCGTAAATTAAGAAAAATGTTTAAAGAGCAGACT	3240
O	3241	CAACCAATATTAATGAAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTTTATAGAA	3300
D	3241	CAACCAATATTAATGAAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTTTATAGAA	3300
O	3301	TAGTTCACGATGAGAAACATTCAGACAGAACTAGTAGAAACAGAGGCCCAAAATGGA	3360
D	3301	TAGTTCACGATGAGAAACATTCAGACAGAACTAGTAGAAACAGAGGCCCAAAATGGA	3360
O	3361	ATGCTATGCTTAGATTAAGGGGTTTTCGAACTGAGGCTTATTAACAAAGTCTTCGGA	3420
D	3361	ATGCTATGCTTAGATTAAGGGGTTTTCGAACTGAGGCTTATTAACAAAGTCTTCGGA	3420
O	3421	GTATTTGTAGCATCTCGAATTAAGAAAGCAAGAAATATGAAGAAAGTACGCTGTA	3480
D	3421	GTATTTGTAGCATCTCGAATTAAGAAAGCAAGAAATATGAAGAAAGTACGCTGTA	3480
O	3481	ATACAGATTTCTCCATATCTGATTCAGATTAAGAACAGCCTATGGAGATGATC	3540
D	3481	ATACAGATTTCTCCATATCTGATTCAGATTAAGAACAGCCTATGGAGATGATC	3540
O	3541	ATGCATCTCAGGTTGTTGTTGAGACACCTGATGACCTGTATGATGATGTTAAAGG	3600
D	3541	ATGCATCTCAGGTTGTTGTTGAGACACCTGATGACCTGTATGATGATGTTAAAGG	3600
O	3601	AAAGTACTAGTTTGGCGAAAAATGACATTAAGGAAATTTCTGCTGTTTAAAGGCG	3660
D	3601	AAAGTACTAGTTTGGCGAAAAATGACATTAAGGAAATTTCTGCTGTTTAAAGGCG	3660
O	3661	TCCAGAAAGAGAGCTAGCAGAGCTCTAGCCCTTTCACCCATATGCTTGGCTCAGG	3720
D	3661	TCCAGAAAGAGAGCTAGCAGAGCTCTAGCCCTTTCACCCATATGCTTGGCTCAGG	3720
O	3721	GTTACCGAAGAGGGGCCAAGAAATTAAGTCTTCAGAGAGAACTTATCTAGTGAGATG	3780
D	3721	GTTACCGAAGAGGGGCCAAGAAATTAAGTCTTCAGAGAGAACTTATCTAGTGAGATG	3780

QY	3781	AAGAGCTCCCTGCTTCCAACTGTTATTGGTAAAGTAACATATACCTTCTAGT	3840	QY	4861	AAGACAGAGCCCGAGAGTACGTCGTGTTGGCAACATACCATTCTCAACCTGCTGATTGA	4920
Db	3781			Db	4861		
QY	3841	CTACTAGGCAATGACACCGTGTGCTACGAGTGTCTGTCTAAGAACACAGAGAGATTAT	3900	QY	4921	AAGTCCCAATTGAAGTTGCGAGATCTGCCAGAGTCCAGCTGCTGCTCACTACTG	4980
Db	3841			Db	4921		
QY	3901	TATCATTTGAAGATGATTAAATGACGTGACAGTAACAGGTAATTTGGCAAGGCAATCTC	3960	QY	4981	ATATGCTGGGTAAATGCAATGGAAGAAAGTGTGACAGGGGAGGAGCCAGAAATGACAG	5040
Db	3901			Db	4981		
QY	3961	AGGACATCACCTTGTAGAGAAACAAATGTTCTGCTAGCTTGTCTTCTCACAGTGA	4020	QY	5041	CTTCACAGAAAGGTCACACAAAGAAATGTCATGTTGTTGCTGCTGACCCAGAAAG	5100
Db	3961			Db	5041		
QY	4021	GTCATTTGGAAGCTTGACGTGCAATATCAACACCCAGATCCTTTCTGATTGTTCTT	4080	QY	5101	AATTTATGCTCGTGTACAAAGTTTGGCAGAAACACCATCTCACTTAACTTAATTA	5160
Db	4021			Db	5101		
QY	4081	CCAAACAATGAGCATCATGTGAAGCCAGGAGTTGGTGTGAGTGAAGGAATGG	4140	QY	5161	CTGAGAGACTACTCATGTTGTTATGAAACACAGATCTGAGTTTGTGTGAACGACAC	5220
Db	4081			Db	5161		
QY	4141	TTTCAGATGATGAAGAAAGAGAAACGGGCTTGGAAAGAAATATCAAGAAAGCAAGCA	4200	QY	5221	TGAATATTTTCTAGCAATTTGCGGAGGAAATGAGTATGACATTTCTGGGTGACCC	5280
Db	4141			Db	5221		
QY	4201	TGATTTCAAACTTAAGTGAAGACACATCTGGGTGTGAGAGTGAACAGGCTCTCTAG	4260	QY	5281	AGTCTATTTAAAGAAAGAAATGCTGAATGAGCATATTTGAAGTCAAGAGAGATGG	5340
Db	4201			Db	5281		
QY	4261	ACGCTCAGGAGCTATCCTCAGAGTGAATTTTAACTCAGCAGAGGATCCATGTC	4320	QY	5341	TCAATGGAAGAAACCCCAAGTGTCAAGAGACAGACAGATCCAGAGAGATGCT	5400
Db	4261			Db	5341		
QY	4321	AACATTAACCTGATTAAGCTCCAGAGAAATGGCTGAAGTGAAGTGTGTTAGAACAGC	4380	QY	5401	TCAGGGGGCTAGAAATCTGTTGCTATGGCCCTTCCACAAATGCCAGATCAACTGG	5460
Db	4321			Db	5401		
QY	4381	ATGGAGCCAGCCTTCTAACAGCTCCCTTCATCATAGTGAATCTTCTGCCCCCTAGG	4440	QY	5461	AATGATGTTGACAGCTGTGTTGCTTCTGTGTGTAAGAGCTTTTCATCTACCCCTTG	5520
Db	4381			Db	5461		
QY	4441	ACCTGGAATCCAGAAACAAAGCAATCAGAAAAGAGTAACTTCACAGAAAGTA	4500	QY	5521	GCACAGGTGTCCACCAATTTGTTGTTGCTGAGAGCTGAGAGAGAGCAATGGT	5580
Db	4441			Db	5521		
QY	4501	GTCATTAACCTTATTAAGCCAGAAATCCAGAAAGGCTTTCTGCTGACAAAGTTGAGGTCTG	4560	QY	5581	TCCATGCAATTTGGGCGAGATGTGTGAGGCACTGTGTGACCCGAGAGTGGTGGACA	5640
Db	4501			Db	5581		
QY	4561	CAGATAGTTCTTACCACTAAATAAAGAACACAGAGTGAAGAGTCAATCCCTTTCTAAT	4620	QY	5641	GTTAGACACTTACCACTGAGAGAGTGTGAGAGCTGAGACCTTACCTGATACCCAGATCCCA	5700
Db	4561			Db	5641		
QY	4621	GCCCATCATTTAGTGTAGTGTGATGACAGTGTGCTGTGGAGTCTTCAGAAATGAA	4680	QY	5701	GCCCACTACTGA	5711
Db	4621			Db	5701		
QY	4681	ACTACCATCTCAAGAGAGGCTATTAAAGTTTGTGATGTGAGAGCAACAGCTGGAAG	4740	QY	5711		
Db	4681			Db	5711		
QY	4741	ACTCTGGGCGACAGATTTGAGCGAAACATCTTACTTCCAGAGCAAGATCTAGAGGAA	4800	QY	5711		
Db	4741			Db	5711		
QY	4801	CCCCTTACTGGAATCTGAATCAGCCTTCTCTATGACCTGGAATCTGATCCTTCTG	4860	QY	5711		
Db	4801			Db	5711		
QY	4861	CCCCTTACTGGAATCTGAATCAGCCTTCTCTATGACCTGGAATCTGATCCTTCTG	4860	QY	5711		
Db	4861			Db	5711		

## RESULT 4

US-09-982-828-3

Sequence 3, Application US/09982828

Publication No. US20030022184A1

## GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.

Alvarres, Christopher P.

Critz, Brenda S.

Olson, Sheri J.

Thurber, Denise

Zeng, Bin

```

: TITLE OF INVENTION: Coding Sequences of the Human
: BRCAL Gene
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Morgan Lewis & Bockius LLP
: STREET: 1111 Pennsylvania Avenue N. W.
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/982,828
: FILING DATE: 22-Oct-2001
: CLASSIFICATION: <Unknown>
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 09/074,453
: FILING DATE: 1998-05-06
: APPLICATION NUMBER: US 08/798,691
: FILING DATE: 1997-02-12
: APPLICATION NUMBER: US 08/598,591
: FILING DATE: 1996-02-12
: ATTORNEY/AGENT INFORMATION:
: NAME: Michael S. Tuscan
: REGISTRATION NUMBER: 43,210
: REFERENCE/DOCKET NUMBER: 44921-5053-01-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-739-3000
: TELEFAX: 202-739-3001
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5711 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: STRAIN: BRCAL (cml12)
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: 17
: MAP POSITION: 17q21
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:
:
: US-09-982-828-3
:
: Query Match          99.9%; Score 5704.6; DB 9; Length 5711;
: Best Local Similarity 99.9%; Pred. No. 0;
: Matches 5707; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 301 GTCCCTTATGTAAGATGATATTAACCAAAAGAGCTACAGAAAGTACGAGATTAGTC 360
: |||||||
Db 301 GTCCCTTATGTAAGATGATATTAACCAAAAGAGAGCTACAGAAAGTACGAGATTAGTC 360
QY 361 AACTTTGTAAGAGCTATTGAAAATCATTTTGCTTTTACGCTTGACACAGGTTTGAGT 420
: |||||||
Db 361 AACTTTGTAAGAGCTATTGAAAATCATTTTGCTTTTACGCTTGACACAGGTTTGAGT 420
QY 421 ATGCAACAGCTATATTTTGCAGAAAAGAAAATTAATCTCTCGTGAACATCTAAAGATG 480
: |||||||
Db 421 ATGCAACAGCTATATTTTGCAGAAAAGAAAATTAATCTCTCGTGAACATCTAAAGATG 480
QY 481 AAGTTTCTATCATCCAAAGATATGGGCTACAGAAAACCGTCCAAAGAGCTTTACAGAGTG 540
: |||||||
Db 481 AAGTTTCTATCATCCAAAGATATGGGCTACAGAAAACCGTCCAAAGAGCTTTACAGAGTG 540
QY 541 AACCCGAAAATCCTTCTCTGAGAGAAACCACTCTCAGTGTCCAACTCTTAACCTTGAA 600
: |||||||
Db 541 AACCCGAAAATCCTTCTCTGAGAGAAACCACTCTCAGTGTCCAACTCTTAACCTTGAA 600
QY 601 CTGTGAGAGCTGTGAGGACAAAGAGGAGATACCAACCTCAAAAGAGCTGTCTACATATG 660
: |||||||
Db 601 CTGTGAGAGCTGTGAGGACAAAGAGGAGATACCAACCTCAAAAGAGCTGTCTACATATG 660
QY 661 AATTGGGATCTGATTTCTTCTGAAGATACCGTTAATAAGGCAACTTATTCAGTGTGGAG 720
: |||||||
Db 661 AATTGGGATCTGATTTCTTCTGAAGATACCGTTAATAAGGCAACTTATTCAGTGTGGAG 720
QY 721 ATCAAGATTTGTTACAAATCACCCCTCAAGSACAGGAGATGAATACGTTTGATTTCTG 780
: |||||||
Db 721 ATCAAGATTTGTTACAAATCACCCCTCAAGSACAGGAGATGAATACGTTTGATTTCTG 780
QY 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAAACAAATCTGAATCATATCAAC 840
: |||||||
Db 781 CAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAAACAAATCTGAATCATATCAAC 840
QY 841 CCAGTAATTAATGATTTGACACCACTGAGAGGCTGACGAGAGGATCCAGAAAAGT 900
: |||||||
Db 841 CCAGTAATTAATGATTTGACACCACTGAGAGGCTGACGAGAGGATCCAGAAAAGT 900
QY 901 ATCAAGGATAGTTCTGTTTCAAACTTGCATGTGAGCCATGTGGCAAAATCTCATGCA 960
: |||||||
Db 901 ATCAAGGATAGTTCTGTTTCAAACTTGCATGTGAGCCATGTGGCAAAATCTCATGCA 960
QY 961 GCTCATTACAGCATGAGACAGCAGTTTATTAATCACTAAAGACAGATGAATGTAGAA 1020
: |||||||
Db 961 GCTCATTACAGCATGAGACAGCAGTTTATTAATCACTAAAGACAGATGAATGTAGAA 1020
QY 1021 AGGCTGAATTTCTGTAATAAAGCAACAGCCTGGCTTAGCAAGGAGCCACATTAACAGAT 1080
: |||||||
Db 1021 AGGCTGAATTTCTGTAATAAAGCAACAGCCTGGCTTAGCAAGGAGCCACATTAACAGAT 1080
QY 1081 GGGCTGGAAGTAAAGAAACATGTATGATAGCGGACTCCAGACAGAAAAAAGGTAG 1140
: |||||||
Db 1081 GGGCTGGAAGTAAAGAAACATGTATGATAGCGGACTCCAGACAGAAAAAAGGTAG 1140
QY 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGAATGAACAGAACTGCATGCT 1200
: |||||||
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAGATGAATGAACAGAACTGCATGCT 1200
QY 1201 CAGAGAAATCCAGAGATACCTGAAGATGTTCTCTTGATTAACACTTAATAGCAATTCAGA 1260
: |||||||
Db 1201 CAGAGAAATCCAGAGATACCTGAAGATGTTCTCTTGATTAACACTTAATAGCAATTCAGA 1260
QY 1261 AAGTTAATGAGTGTTCAGAGAGTATGATGATAGTTAGTTGATGACTACATGATG 1320
: |||||||
Db 1261 AAGTTAATGAGTGTTCAGAGAGTATGATGATAGTTAGTTGATGACTACATGATG 1320
QY 1321 GGGAGCTGGAATCAAAAGCCAAAGTACGCTGATGATGATGAGCTTAATAGGTAGATG 1380
: |||||||
Db 1321 GGGAGCTGGAATCAAAAGCCAAAGTACGCTGATGATGATGAGCTTAATAGGTAGATG 1380
QY 1381 AATATCTGCTTCTTCTGAGAGAAATAGACTTACGAGCATGATCTCATAGGCTTTAA 1440
: |||||||

```

```

1381 AATATTCTGGTCTTCAGAGAAAATAGACTTACTGGCCACTGATCCCATGAGGCTTTAA 1440
1441 TATGTAAAGTGAAGAGTTCCTCCAAATCAGTAGAGTAATATTTGAAGACAAATAT 1500
1441 TATGTAAAGTGAAGAGTTCCTCCAAATCAGTAGAGTAATATTTGAAGACAAATAT 1500
1501 TTGGGAAAACCTATCGGAAGAAAGCAAGCCTCCCACTTAAGCCATGTAAGTGAATAT 1560
1501 TTGGGAAAACCTATCGGAAGAAAGCAAGCCTCCCACTTAAGCCATGTAAGTGAATAT 1560
1561 TTAATTATAGAGCATTGTTTACTGAGCCACAGATPAATACAAAGCGTCCCTCACAATA 1620
1561 TTAATTATAGAGCATTGTTTACTGAGCCACAGATPAATACAAAGCGTCCCTCACAATA 1620
1621 AATTTAAGCGTAAAGAGACCTACATCAGGCGCTTCACTCGAGATTATATCAAGAAAG 1680
1621 AATTTAAGCGTAAAGAGACCTACATCAGGCGCTTCACTCGAGATTATATCAAGAAAG 1680
1681 CAGATTTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAACCTAACCAACGAGAC 1740
1681 CAGATTTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAACCTAACCAACGAGAC 1740
1741 AAGAAATGCTAAGATGATATTTACTAATAGTGTCTCATGAGATTAACCAAAAGGTGATT 1800
1741 AAGAAATGCTAAGATGATATTTACTAATAGTGTCTCATGAGATTAACCAAAAGGTGATT 1800
1801 CTATTTCAGATGAGAAAATCTTAACCCATAGATCAGCAAAAAGAAATCTGCTTTCA 1860
1801 CTATTTCAGATGAGAAAATCTTAACCCATAGATCAGCAAAAAGAAATCTGCTTTCA 1860
1861 AAACGAAAGCTGAACCTATTAAGCAGAGATTAAGCAATATGAACTCGAATTAATATCC 1920
1861 AAACGAAAGCTGAACCTATTAAGCAGAGATTAAGCAATATGAACTCGAATTAATATCC 1920
1921 ACAAATTTCAAAGACCTTAAAGAAATAGGCTGAGAGGAAAGTCTTACACGACATATTC 1980
1921 ACAAATTTCAAAGACCTTAAAGAAATAGGCTGAGAGGAAAGTCTTACACGACATATTC 1980
1981 ATGCGCTTGAATAGTACAGTACAGTAAATCTTAAGCCACCTAATTTAGTAATGCGAAA 2040
1981 ATGCGCTTGAATAGTACAGTACAGTAAATCTTAAGCCACCTAATTTAGTAATGCGAAA 2040
2041 TTGATATGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACCAACCAATGCCAGTCA 2100
2041 TTGATATGTTGTTCTAGCAGTGAAGATTAAGAAAAAGTACCAACCAATGCCAGTCA 2100
2101 GGACACAGCAGAAACCTACAACTCATGGAAGGTAAAGACCTGCACTGGAGCCAGAGAA 2160
2101 GGACACAGCAGAAACCTACAACTCATGGAAGGTAAAGACCTGCACTGGAGCCAGAGAA 2160
2161 GTTACAAAGCCAATGAAAGACAGACAAAGTAAAGACATGACATGACTTTCCAGAGCTGA 2220
2161 GTTACAAAGCCAATGAAAGACAGACAAAGTAAAGACATGACATGACTTTCCAGAGCTGA 2220
2221 AGTTAAACAATGACACCTGGTCTTTTACTTAAGTGTCAATACCAATGCAACTTAAGAAAT 2280
2221 AGTTAAACAATGACACCTGGTCTTTTACTTAAGTGTCAATACCAATGCAACTTAAGAAAT 2280
2281 TTGTCAATCTAGCCCTTCCAAGAGAAAGAAAGAAAGCAAGAAACAGTTAAAGCTG 2340
2281 TTGTCAATCTAGCCCTTCCAAGAGAAAGAAAGAAAGCAAGAAACAGTTAAAGCTG 2340
2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCAGAACTG 2400
2341 CTAAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCAGAACTG 2400
2401 AAAGATCTGTAGAGAGTACAGTATTTTCACTGCTGACTGTATATATGGAACATCAGG 2460
2401 AAAGATCTGTAGAGAGTACAGTATTTTCACTGCTGACTGTATATATGGAACATCAGG 2460
2461 AAAGATCTGCTTACTGGAAGTTAGCAGTCTAGGGAAGCAAAACAGAAACCAATTAAT 2520
2461 AAAGATCTGCTTACTGGAAGTTAGCAGTCTAGGGAAGCAAAACAGAAACCAATTAAT 2520

```

OY	3601	AAGATCTAGTTCCTGCTCAAAAATGCACATTAAAGAAAGTTCTGTCTTTTAAAGCAAACCG	3660
Db	3601	AAGATCTAGTTCCTGCTCAAAAATGCACATTAAAGAAAGTTCTGTCTTTTAAAGCAAACCG	3660
OY	3661	TCCAGAAAGSAGAGGTTTGACGAGAGTCTCACCCTTCACCATAACATATTGGCTCAGG	3720
Db	3661	TCCAGAGAGGAGAGCTTAGCAGAGTCTCACCCTTCACCATAACATATTGGCTCAGG	3720
OY	3721	GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAGAGACTTATCTAGTAGAGATG	3780
Db	3721	GTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAGAGACTTATCTAGTAGAGATG	3780
OY	3781	AAGAGCTTCCTGCTTCCAACACTTGTATTGGTAAGTAACAATATACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCTGCTTCCAACACTTGTATTGGTAAGTAACAATATACCTTCTCAGT	3840
OY	3841	CTACTAGCATAGCACCGCTGCTCAACCGAGTCTGCTCAACACACAGAGGAATTTAT	3900
Db	3841	CTACTAGCATAGCACCGCTGCTCAACCGAGTCTGCTCAACACACAGAGGAATTTAT	3900
OY	3901	TATCATTTGAAGAAATAGCTTTAAATGACTGCAGTAACACAGTAAATTTGGCAAAGCATCTC	3960
Db	3901	TATCATTTGAAGAAATAGCTTTAAATGACTGCAGTAACACAGTAAATTTGGCAAAGCATCTC	3960
OY	3961	AGGAACATCACCTTAGTAGGAGAACAAATTTCTGCTAGTGTGTTTCTTCACAGTCA	4020
Db	3961	AGGACATCACCTTAGTAGGAGAACAAATTTCTGCTAGTGTGTTTCTTCACAGTCA	4020
OY	4021	GTTAAATTGGAAAGACTTGACTCAATATCAACACCCAGAGTCTTTCTTGATTTGGTTCTT	4080
Db	4021	GTTAAATTGGAAAGACTTGACTCAATATCAACACCCAGAGTCTTTCTTGATTTGGTTCTT	4080
OY	4081	CCAAACAATAGSGCATAGTCTGAACACCAGGAGTTGGTCTGAGTACAAAGAAATTGG	4140
Db	4081	CCAAACAATAGSGCATAGTCTGAACACCAGGAGTTGGTCTGAGTACAAAGAAATTGG	4140
OY	4141	TTTTGAGATGAGAAAGAAAGAGAACGGGCTTGGAAAGAAATATCAAGAAAGCAAAGCA	4200
Db	4141	TTTTGAGATGAGAAAGAAAGAGAACGGGCTTGGAAAGAAATATCAAGAAAGCAAAGCA	4200
OY	4201	TGGATTTCAAACCTTAGGTGAAGCAGACATCTGGGTGTGAAGTGAACAAAGCCTCTCTGAG	4260
Db	4201	TGGATTTCAAACCTTAGGTGAAGCAGACATCTGGGTGTGAAGTGAACAAAGCCTCTCTGAG	4260
OY	4261	ACTGCTCAGGGCTATCCTCTCAGAGTACATTTTAAACCACCTCAGAGAGGGATACCATGC	4320
Db	4261	ACTGCTCAGGGCTATCCTCTCAGAGTACATTTTAAACCACCTCAGAGAGGGATACCATGC	4320
OY	4321	AACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAACCTGAAGCTGTGTAGAACAGC	4380
Db	4321	AACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAACCTGAAGCTGTGTAGAACAGC	4380
OY	4381	ATGGGAGCAGCCTTTAAACAGCTTAACCTTCATCATTAAGTAGACTCTTCTCCCTTGAGG	4440
Db	4381	ATGGGAGCAGCCTTTAAACAGCTTAACCTTCATCATTAAGTAGACTCTTCTCCCTTGAGG	4440
OY	4441	ACCTGCGAAAAATCCAGAACAAAGACATCAGAAAAAGCAGTTTAACCTTCACAGAAAAAGTA	4500
Db	4441	ACCTGCGAAAAATCCAGAACAAAGACATCAGAAAAAGCAGTTTAACCTTCACAGAAAAAGTA	4500
OY	4501	GTGAATACCTCTATTAAGCCAGATCCAGAGGCTTTTCTGCTGACAGTTTGGAGTGTCTG	4560
Db	4501	GTGAATACCTCTATTAAGCCAGATCCAGAGGCTTTTCTGCTGACAGTTTGGAGTGTCTG	4560
OY	4561	CAGATAGTTCACAGTAAATAATTAAGAACCCAGAGTGGAAAGTTCATCCCCTTCTTAAT	4620
Db	4561	CAGATAGTTCACAGTAAATAATTAAGAACCCAGAGTGGAAAGTTCATCCCCTTCTTAAT	4620
OY	4621	GCCCATCATTAAGATGATAGTGTGTRCATGCACAGTTGCTGGAGAGTCTTCAGAAATGAA	4680
Db	4621	GCCCATCATTAAGATGATAGTGTGTRCATGCACAGTTGCTGGAGAGTCTTCAGAAATGAA	4680

QY	4661	ACTACCCATCTCAGAGAGCTCATTAAGTTGTTGATGTGGAGAGCACAGCTGGAA	4740
Db	4661	ACTACCCATCTCAGAGAGAGCTCATTAAGTTGTTGATGTGGAGAGCACAGCTGGAA	4740
QY	4741	AGTCGGGCGCCACAGATTGTGACGGAAACATCTTACTTCCCAAGCAGAGATCTACAGGAA	4800
Db	4741	AGTCGGGCGCCACAGATTGTGACGGAAACATCTTACTTCCCAAGCAGAGATCTACAGGAA	4800
QY	4801	CCCCCTTACCTGGAAATCTGGAATCA6CCCTTCTCTGATGACCCGTAATCTGATCCCTTCTG	4860
Db	4801	CCCCCTTACCTGGAAATCTGGAATCA6CCCTTCTCTGATGACCCGTAATCTGATCCCTTCTG	4860
QY	4861	AAGACAGAGCCCGCAGAGTCACTGCTGTGTGGCAACATACCACTTCTCAACCTCTGCATTGGA	4920
Db	4861	AAGACAGAGCCCGCAGAGTCACTGCTGTGTGGCAACATACCACTTCTCAACCTCTGCATTGGA	4920
QY	4921	AAGTTCGCCCAATTTGAAAGTTGTCAGAAATGTGCGCAGAGTCCAGCTGCTGCATACCTCTG	4980
Db	4921	AAGTTCGCCCAATTTGAAAGTTGTCAGAAATGTGCGCAGAGTCCAGCTGCTGCATACCTCTG	4980
QY	4981	ATACTGCTGGCTATTAATGCAATGGAAGAAATGTGAGCAGGAGAGAGCCAGAAATTTGACAG	5040
Db	4981	ATACTGCTGGCTATTAATGCAATGGAAGAAATGTGAGCAGGAGAGAGCCAGAAATTTGACAG	5040
QY	5041	CTTCACAGAAAGGCTTCACAAAAGAAATGTCATGAGTGTGCTGCGGCTGACCCCAAG	5100
Db	5041	CTTCACAGAAAGGCTTCACAAAAGAAATGTCATGAGTGTGCTGCGGCTGACCCCAAG	5100
QY	5101	AATTTATGCTGCTGTCAGAGTTTCCAGAAAACACACATCACTTAACTAATCTAATTA	5160
Db	5101	AATTTATGCTGCTGTCAGAGTTTCCAGAAAACACACATCACTTAACTAATCTAATTA	5160
QY	5161	CTGAAGAGACATACATCTGTTGTAAGAAACAGATGCTGAGTTGTGTGTGAACAGGACAC	5220
Db	5161	CTGAAGAGACATACATCTGTTGTAAGAAACAGATGCTGAGTTGTGTGTGAACAGGACAC	5220
QY	5221	TGAAATATTTTCTAGGAATTCGGGAGAGAAATGGTAGTTAGCTATTTTCTGGGTGACCC	5280
Db	5221	TGAAATATTTTCTAGGAATTCGGGAGAGAAATGGTAGTTAGCTATTTTCTGGGTGACCC	5280
QY	5281	AGCTATTTAAAGAAAGAAATTCGTGAATGACATGATTTTGAAGTCAAGAGACATGTGG	5340
Db	5281	AGCTATTTAAAGAAAGAAATTCGTGAATGACATGATTTTGAAGTCAAGAGACATGTGG	5340
QY	5341	TCAATGGAAGAAACACCAAGCTCCAAAGCAGAGAGAGAAATCCCAAGACAGAAAGATCT	5400
Db	5341	TCAATGGAAGAAACACCAAGCTCCAAAGCAGAGAGAGAAATCCCAAGACAGAAAGATCT	5400
QY	5401	TCAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACACCAACATGCCCCACAGATCACTGG	5460
Db	5401	TCAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACACCAACATGCCCCACAGATCACTGG	5460
QY	5461	AATGATGCTGACACTGTGTGCTGCTTCTGTGTGAAGAGACTTTCATCATATCACCTTG	5520
Db	5461	AATGATGCTGACACTGTGTGCTGCTTCTGTGTGAAGAGACTTTCATCATATCACCTTG	5520
QY	5521	GCACAGAGTGTCCACCAATTTGCTTGTGACGACAGATGCTGTGACAGAGACAAATGGCT	5580
Db	5521	GCACAGAGTGTCCACCAATTTGCTTGTGACGACAGATGCTGTGACAGAGACAAATGGCT	5580
QY	5581	TTCATGCAATTTGGGCAGATGTGTGACGACCTGTGTGTGACCCGAGAGTGGGTGTTGGACA	5640
Db	5581	TTCATGCAATTTGGGCAGATGTGTGACGACCTGTGTGTGACCCGAGAGTGGGTGTTGGACA	5640
QY	5641	GTTGAGCACTTACACAGTCCAGAGGCTGGACACCTTACCTGATATCCCCACATCCCCACA	5700
Db	5641	GTTGAGCACTTACACAGTCCAGAGGCTGGACACCTTACCTGATATCCCCACATCCCCACA	5700
QY	5701	GCCACATCACTGA 5711	
Db	5701	GCCACATCACTGA 5711	

```

RESULT 5
US-10-022-819-1
: Sequence 1, Application US/10022819
: Publication No. US20030027166A1
: GENERAL INFORMATION:
: APPLICANT: ALLEN, Antonette C. P.
: OLESEN, Sheri J.
: LAWRENCE, Tammy
: ANGELLY, Tracy S.
: RABIN, Mark B.
: TITLE OF INVENTION: CODING SEQUENCE HAPLOTYPE OF THE HUMAN
: BRCAL GENE
: NUMBER OF SEQUENCES: 67
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Morgan Lewis & Bockius LLP
: STREET: 1111 Pennsylvania Avenue
: CITY: Washington DC
: STATE: District of Columbia
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/022,819
: FILING DATE: 22-Apr-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/074,452
: FILING DATE: 1998-05-06
: ATTORNEY/AGENT INFORMATION:
: NAME: <Unknown>
: REGISTRATION NUMBER: <Unknown>
: REFERENCE/DOCKET NUMBER: 044921-5049-01-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-739-3000
: TELEFAX: 202-739-3001
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5711 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: HOMO SAPIENS
: STRAIN: BRCAL
: HAPLOTYPE: OM14
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: 17
: MAP POSITION: 17q21
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-022-819-1

Query Match          99.9%: Score 5704.6; DB 9; Length 5711;
Best Local Similarity 99.9%: Pred. No. 0;
Matches 5707: Conservative 0; Mismatches 4; Indels 0; Gaps 0.

QY      1 AGCTGCGTGAAGACTCTGTGAACCCCGACACAGAGCTGTGGGGTTTCTCAGATPACTGGGCC 60
        |||||||
Db       1 AGCTGCGTGAAGACTCTGTGAACCCCGACACAGAGCTGTGGGGTTTCTCAGATPACTGGGCC 60

QY      61 CCTGCGCTCAGAGAGCCCTTCAACCCCTCTGCTCTGGGTTAAAGTTCAATTGGAACGAAAGAAA 120
        |||||||
Db       61 CCTGCGCTCAGAGAGCCCTTCAACCCCTCTGCTCTGGGTTAAAGTTCAATTGGAACGAAAGAAA 120

QY      121 TGGATTTATCTGCTCTTGGCGGTGGAAGAAGTACAAATGTCATTATCTTATGCAGAAA 180
        |||||||
Db       121 TGGATTTATCTGCTCTTGGCGGTGGAAGAAGTACAAATGTCATTATCTTATGCAGAAA 180

```

Db	121	TGGATTATATCGCTCTTCCGGTTTGAAGAAGTACAAAAATGTGATTAAATGCTATATGCTATGACAGAAA	180
QY	181	TCCTTAGAGTGTCCCAATCTGTCTGAGACTTGATCAAGGAACGTGTCTCCAGAAAGTGTACCC	240
Db	181	TCCTTAGAGTGTCCCAATCTGTCTGAGACTTGATCAAGGAACGTGTCTCCAGAAAGTGTACCC	240
QY	241	ACATATTTTGCANNTTTTGCATGCTGGAACCTTCACACACAAAGAGGCTTTCACAGT	300
Db	241	ACATATTTTGCANNTTTTGCATGCTGGAACCTTCACACACAAAGAGGCTTTCACAGT	300
QY	301	GTCTTATATGTAAGATGATATACCAAGAGGCTCTCAAGAAAGTACGATATTTAGTC	360
Db	301	GTCTTATATGTAAGATGATATACCAAGAGGCTCTCAAGAAAGTACGATATTTAGTC	360
QY	361	AACCTGTTGAAGAGCTATTGAAATCAATTTGTCTTTTCAGCTTGACACAGGTTTGAGAT	420
Db	361	AACCTGTTGAAGAGCTATTGAAATCAATTTGTCTTTTCAGCTTGACACAGGTTTGAGAT	420
QY	421	ATGCAAAACACTTAATTTTGCAGAAAAAGGAAATTAATCTTCCTGCAACATCTTAAAGATG	480
Db	421	ATGCAAAACACTTAATTTTGCAGAAAAAGGAAATTAATCTTCCTGCAACATCTTAAAGATG	480
QY	481	AAGTTTCTATCATCCAAAGATGAGGCTACAGAAACCGTGCCAAAGACTCTACAGAGTG	540
Db	481	AAGTTTCTATCATCCAAAGATGAGGCTACAGAAACCGTGCCAAAGACTCTACAGAGTG	540
QY	541	AACCCGAAAAATCCCTTCCTTCGACAGGAACACAGTCTCAGTCCAACTCTTAACCTTGGAA	600
Db	541	AACCCGAAAAATCCCTTCCTTCGACAGGAACACAGTCTCAGTCCAACTCTTAACCTTGGAA	600
QY	601	CTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACTCAAGAGAGCTCTGTCTACATTTG	660
Db	601	CTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACTCAAGAGAGCTCTGTCTACATTTG	660
QY	661	AATTGGATCTGATTTCTCTGAAAGATACCGTTAATAAGGCAACTTATTGAGTGGGAG	720
Db	661	AATTGGATCTGATTTCTCTGAAAGATACCGTTAATAAGGCAACTTATTGAGTGGGAG	720
QY	721	ATCAGAAATTTGTACCAATCACCCCTCAAGGACACGAGGATGAATAGTTTGATTCG	780
Db	721	ATCAGAAATTTGTACCAATCACCCCTCAAGGACACGAGGATGAATAGTTTGATTCG	780
QY	781	CAAAAAAGGCTGTTGTGAATTTTCTGAGACGGATGTACAAATACTGAACATCATCAAC	840
Db	781	CAAAAAAGGCTGTTGTGAATTTTCTGAGACGGATGTACAAATACTGAACATCATCAAC	840
QY	841	CCAGTAATATATGATTTGGAACACCACTGGAAGCGTGACGTCGAGAGGCAATCCGAAAAAT	900
Db	841	CCAGTAATATATGATTTGGAACACCACTGGAAGCGTGACGTCGAGAGGCAATCCGAAAAAT	900
QY	901	ATAGAGGATGTTCTGTTTCAAACTTGCATGTGAGACATGTGGCACAAATACATGCGCA	960
Db	901	ATAGAGGATGTTCTGTTTCAAACTTGCATGTGAGACATGTGGCACAAATACATGCGCA	960
QY	961	GCTCATTTACAGCATGAGAACAGCAGTATTACTACTATAAGACAGATGATGTAGANA	1020
Db	961	GCTCATTTACAGCATGAGAACAGCAGTATTACTACTATAAGACAGATGATGTAGANA	1020
QY	1021	AGGCTGAATTTCTGTAATAAAGCAAAACGCTTGACTTAGCAAGAGGCAACATTAACAGAT	1080
Db	1021	AGGCTGAATTTCTGTAATAAAGCAAAACGCTTGACTTAGCAAGAGGCAACATTAACAGAT	1080
QY	1081	GGGCTGGAAGTAAAGAACATGATATGATAGCGGACCTCCAGACACAGAAAAAAGCTGAG	1140
Db	1081	GGGCTGGAAGTAAAGAACATGATATGATAGCGGACCTCCAGACACAGAAAAAAGCTGAG	1140
QY	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGGAATTAAGCAGAAATCTGCCATGCT	1200
Db	1141	ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGGAATTAAGCAGAAATCTGCCATGCT	1200
QY	1201	CAGAGATCTGAGAGTACTGAAGATGTTCTTGAGTATACACTTAATAGCAGCATTCAGA	1260
Db	1201	CAGAGATCTCTGAGAGTACTGAAGATGTTCTTGAGTATACACTTAATAGCAGCATTCAGA	1260

OY	1261	AAGTTAAATGAGTGGTTTCCAGAGATGATGAACCTGTTAGTGTTCGATGACTCACATGATG	1320
Db	1261	AAGTTAAATGAGTGGTTTCCAGAGATGATGAACCTGTTAGTGTTCGATGACTCACATGATG	1320
OY	1321	GGGAGTCGAAATCAAAATGCCAAACTAGCTATGATATGGAGCTCTAAATGAGGTAAATG	1380
Db	1321	GGGAGTCGAAATCAAAATGCCAAACTAGCTATGATATGGAGCTCTAAATGAGGTAAATG	1380
OY	1381	AATATTCGTGTTCTTCAGAGAAATAATACCTACTAGGCGAGTGAATCCATGAGGCTTTAA	1440
Db	1381	AATATTCGTGTTCTTCAGAGAAATAATACCTACTAGGCGAGTGAATCCATGAGGCTTTAA	1440
OY	1441	TATGTAAAAGTGAAGAAGCTTCATCCAAATCAGTAGAGATATATTTGAAGACAAATAT	1500
Db	1441	TATGTAAAAGTGAAGAAGCTTCATCCAAATCAGTAGAGATATATTTGAAGACAAATAT	1500
OY	1501	TTGGGAAAACCTATCGGAGAGAAGGCAAGCCCTCCCACTTAAGGCATGTAACTGAATTC	1560
Db	1501	TTGGGAAAACCTATCGGAGAGAAGGCAAGCCCTCCCACTTAAGGCATGTAACTGAATTC	1560
OY	1561	TAAATTATAGGACATTTTGTTACTGAGCCACAGATAAATACAAAGACGTCCTCCACAATA	1620
Db	1561	TAAATTATAGGACATTTTGTTACTGAGCCACAGATAAATACAAAGACGTCCTCCACAATA	1620
OY	1621	AATTAAGAGCGTAAAGAGAGACCTCATCAGAGCCCTCATCCGAGAGATTTTATCAAGAAAG	1680
Db	1621	AATTAAGAGCGTAAAGAGAGACCTCATCAGAGCCCTCATCCGAGAGATTTTATCAAGAAAG	1680
OY	1681	CAGATTTTGGCAGTTCAAAAAGACCTCCTGAATGATAAATCAAGGGAACCTAACCMAAGGAGC	1740
Db	1681	CAGATTTTGGCAGTTCAAAAAGACCTCCTGAATGATAAATCAAGGGAACCTAACCMAAGGAGC	1740
OY	1741	AGATGGTCGAAGTATGAATTTTCTTAATATAGTGTCAATGAGATTAATAAACAAGGTATTT	1800
Db	1741	AGATGGTCGAAGTATGAATTTTCTTAATATAGTGTCAATGAGATTAATAAACAAGGTATTT	1800
OY	1801	CTATTTCGAATGAGAAAAAATCTTAACCCAAATAGATATCTCGAAAAAAGAAATCTGCTTTCA	1860
Db	1801	CTATTTCGAATGAGAAAAAATCTTAACCCAAATAGATATCTCGAAAAAAGAAATCTGCTTTCA	1860
OY	1861	AAACGAAAGCTGAACCTTAATAAGCAGCAGTATAGCAATATGGAACCTGGAATTAATATTC	1920
Db	1861	AAACGAAAGCTGAACCTTAATAAGCAGCAGTATAGCAATATGGAACCTGGAATTAATATTC	1920
OY	1921	ACAAATTCAAAAGCAGCTTAATAAAGATAGGCGTGGAGAGAGAGTCTTTCACAGGCATATTC	1980
Db	1921	ACAAATTCAAAAGCAGCTTAATAAAGATAGGCGTGGAGAGAGAGTCTTTCACAGGCATATTC	1980
OY	1981	ATGCGCTTGAACCTAGTAGTACGTAGAAATCTAAGCCACCTAATTTGTAAGTATGGCAAA	2040
Db	1981	ATGCGCTTGAACCTAGTAGTACGTAGAAATCTAAGCCACCTAATTTGTAAGTATGGCAAA	2040
OY	2041	TTGATATGTTGTTTCAGAGTGAAGTAAGATTAAGAAAAAAGTAAACCAATAGCCAGCTCA	2100
Db	2041	TTGATATGTTGTTTCAGAGTGAAGTAAGATTAAGAAAAAAGTAAACCAATAGCCAGCTCA	2100
OY	2101	GGCAGACAGAAACCTCAACACTCATGGAAGTAAAGAACCTGCAACTGAGACCAAGAGA	2160
Db	2101	GGCAGACAGAAACCTCAACACTCATGGAAGTAAAGAACCTGCAACTGAGACCAAGAGA	2160
OY	2161	GTAAACAAGCCAAATGAACAGACAGATTAAGAAAGCATGACAGTATCTTTCCCAAGCTGA	2220
Db	2161	GTAAACAAGCCAAATGAACAGACAGATTAAGAAAGCATGACAGTATCTTTCCCAAGCTGA	2220
OY	2221	AGTTAACAATGCAACCTGCTTTCTTTACTAAGTCTTCAAAATACACAGTAACTTAAAGAT	2280
Db	2221	AGTTAACAATGCAACCTGCTTTCTTTACTAAGTCTTCAAAATACACAGTAACTTAAAGAT	2280
OY	2281	TTTGTCATCTAGCCTTCCAAAGAGAGAAAAAGAGAAACTTGAAGAACGTTAAAGTGT	2340
Db	2281	TTTGTCATCTAGCCTTCCAAAGAGAGAAAAAGAGAAACTTGAAGAACGTTAAAGTGT	2340

QY	2341	CTAATATATGCTGGAGAGACCCCAAAGATCTCATGTTAAAGTGGAGAAAGGGTTTGGAACTG	2400
Db	2341	CTAAATAATGCTGGAGAGACCCCAAAGATCTCATGTTAAAGTGGAGAAAGGGTTTGGAACTG	2400
QY	2401	AAAGTCTGTAGAGATGAGAGTATTTTCATTGGTACCGGATGATTAATGATGGCATCAGG	2460
Db	2401	AAAGATCTGTAGAGATGAGAGTATTTTCATTGGTACCGGATGATTAATGATGGCATCAGG	2460
QY	2461	AAAGTATCTGTTACTGGAGATTTAGACATCTAGGGAAGGCGAAAAACAGAACCAATTAAT	2520
Db	2461	AAAGTATCTGTTACTGGAGATTTAGACATCTAGGGAAGGCGAAAAACAGAACCAATTAAT	2520
QY	2521	GTGTGAGTCAGTGTGACAGCATTTGAAAAACCCCAAGGACTAATTCATGTGTTCCAAAG	2580
Db	2521	GTGTGAGTCAGTGTGACAGCATTTGAAAAACCCCAAGGACTAATTCATGTGTTCCAAAG	2580
QY	2591	ATATATAGAAATAGACACGAGAGGCTTTAATGATTCATTGGGAGCATGAAGTTAACCACAGTC	2640
Db	2591	ATATATAGAAATAGACACGAGAGGCTTTAATGATTCATTGGGAGCATGAAGTTAACCACAGTC	2640
QY	2641	GGGAACAAGCATAGAAATGGAGAGAAAGTGAATGTGATGCTCAGTATTTGGCAGATACAT	2700
Db	2641	GGGAACAAGCATAGAAATGGAGAGAAAGTGAATGTGATGCTCAGTATTTGGCAGATACAT	2700
QY	2701	TCAGGTTTCAAAGCGCCAGTCATTTGCTGTGTTTCAAATCCAGAAATGCAAGAGG	2760
Db	2701	TCAGGTTTCAAAGCGCCAGTCATTTGCTGTGTTTCAAATCCAGAAATGCAAGAGG	2760
QY	2761	AATGTGCAACATTTCTCTGCCACCTCTGGGTCTCTTAAGAAACAAAGTCCAAATCTCACTT	2820
Db	2761	AATGTGCAACATTTCTCTGCCACCTCTGGGTCTCTTAAGAAACAAAGTCCAAATCTCACTT	2820
QY	2821	TTGAATGTGAACAAAAGAGAAAGAAATCAAGSAAAGATGACTTAATATCAAGCTGTAC	2880
Db	2821	TTGAATGTGAACAAAAGAGAAAGAAATCAAGSAAAGATGACTTAATATCAAGCTGTAC	2880
QY	2881	AGACAGTTAATATCACTGACAGGCTTTCCTGTGGTTGGTCAGAAAGATTAAGCCACTTGATA	2940
Db	2881	AGACAGTTAATATCACTGACAGGCTTTCCTGTGGTTGGTCAGAAAGATTAAGCCACTTGATA	2940
QY	2941	ATGCCAATGTATGATTCAAAGSAGGCTTACGTTTGTCTATCATCTTCAGTTCCAGAGCA	3000
Db	2941	ATGCCAATGTATGATTCAAAGSAGGCTTACGTTTGTCTATCATCTTCAGTTCCAGAGCA	3000
QY	3001	ACGAACCTGGAGCTATTCTCCAAATBAACATGGACCTTTTCAAAACCCCATATGCTATAC	3060
Db	3001	ACGAACCTGGAGCTATTCTCCAAATBAACATGGACCTTTTCAAAACCCCATATGCTATAC	3060
QY	3061	CACCACTTTTCCCATCAAGTCATTTTGTTTAAACATAATGTAAGAAAAATCTGCTAGAGG	3120
Db	3061	CACCACTTTTCCCATCAAGTCATTTTGTTTAAACATAATGTAAGAAAAATCTGCTAGAGG	3120
QY	3121	AAAACCTTGGAGAACATTCATATGTCACCTGAAAGAGAAATGGGAATGAGAACCTTCCAA	3180
Db	3121	AAAACCTTGGAGAACATTCATATGTCACCTGAAAGAGAAATGGGAATGAGAACCTTCCAA	3180
QY	3181	GTACAGTGAACAAATTTAGCCGTATBAACATTAGAGAAAATGTTTTAAAGAACCCAGCT	3240
Db	3181	GTACAGTGAACAAATTTAGCCGTATBAACATTAGAGAAAATGTTTTAAAGAACCCAGCT	3240
QY	3241	CAAGCAATATTTAATGAAGTAGTTCACAGTACTAATGAAGTGGGCTCCAGTATTAATGAAA	3300
Db	3241	CAAGCAATATTTAATGAAGTAGTTCACAGTACTAATGAAGTGGGCTCCAGTATTAATGAAA	3300
QY	3301	TAGGTTCCAGTATGAAACATTTCAAGCAGAACTAGGTAAAGAAACAGAGGCGCAAAATTTGA	3360
Db	3301	TAGGTTCCAGTATGAAACATTTCAAGCAGAACTAGGTAAAGAAACAGAGGCGCAAAATTTGA	3360
QY	3361	ATGCTATGCTTATGATTAAGGCTTTTTCGACCTGAGGCTCTATTAACCAAAAGTCTTCCTGGAA	3420
Db	3361	ATGCTATGCTTATGATTAAGGCTTTTTCGACCTGAGGCTCTATTAACCAAAAGTCTTCCTGGAA	3420
QY	3421	GTAATTTGTAGCATCTCGAATAATAAAAAAGCAAGAATATGAAGAAGTGAAGTTCAGACTGTTA	3480



Db	3421	GTAAATGTAAGCATCCGGAATATAAAAGAAAGATATGAAGAAGTTCAGACTCTTA	3480
Oy	3481	ATACAGATTTCTCTCATATCTGATTTACATTAACCTATAGAACACCTATGGAGTATGTC	3540
Db	3481	ATACAGATTTCTCTCATATCTGATTTACATTAACCTATAGAACACCTATGGAGTATGTC	3540
Oy	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATAAAG	3600
Db	3541	ATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATAAAG	3600
Oy	3601	AAGATCTAGTGTGTAAGAAATGACATTAAGAAAGTTCGCTGTTTTAGCAAAAGCG	3660
Db	3601	AAGATCTAGTGTGTAAGAAATGACATTAAGAAAGTTCGCTGTTTTAGCAAAAGCG	3660
Oy	3661	TCCAGAAAGAGAGCTTACAGAGAGTCTAGCCCTTACCCATACATTTGGCTCAGG	3720
Db	3661	TCCAGAGAGAGAGCTTACAGAGAGTCTAGCCCTTACCCATACATTTGGCTCAGG	3720
Oy	3721	GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCTCAGAGAACTTATCTAGTAGAGATG	3780
Db	3721	GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCTCAGAGAACTTATCTAGTAGAGATG	3780
Oy	3781	AAGAGCTTCCCTGCTTCCAACTTGTATTTGTTAAGTAACAAATATACCTTTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTTCCAACTTGTATTTGTTAAGTAACAAATATACCTTTCAGT	3840
Oy	3841	CTACTAGGCATAGCACCGTTGCTACCGAGTGTCTCTTAAGAACACAGAGAGAAATTTAT	3900
Db	3841	CTACTAGGCATAGCACCGTTGCTACCGAGTGTCTCTTAAGAACACAGAGAGAAATTTAT	3900
Oy	3901	TATCATTTGAAGATTAAGCTTAATGACTGCAAGTAAATATTTGGCAAGGCATCTC	3960
Db	3901	TATCATTTGAAGATTAAGCTTAATGACTGCAAGTAAATATTTGGCAAGGCATCTC	3960
Oy	3961	AGCAACATACCTTAGTGTAGGAACAACAATGTTCTAGCTGTCTTCTTCACATGCA	4020
Db	3961	AGCAACATACCTTAGTGTAGGAACAACAATGTTCTAGCTGTCTTCTTCACATGCA	4020
Oy	4021	GTGAATTTGGAGACTTGACATGCAAAATCAAAACCCAGATCCCTTCTGATTTGTTCTT	4080
Db	4021	GTGAATTTGGAGACTTGACATGCAAAATCAAAACCCAGATCCCTTCTGATTTGTTCTT	4080
Oy	4081	CCAAACAAATGAGCATCATGCTGTAAGCCAGGAGTGTCTGAGTGACAGAGAAATGG	4140
Db	4081	CCAAACAAATGAGCATCATGCTGTAAGCCAGGAGTGTCTGAGTGACAGAGAAATGG	4140
Oy	4141	TTTCAGATGATGAAGAAAGAGAGCGGCTTGGAAAGAAATATCAAGAAAGCAAGCA	4200
Db	4141	TTTCAGATGATGAAGAAAGAGAGCGGCTTGGAAAGAAATATCAAGAAAGCAAGCA	4200
Oy	4201	TGGATTTCAAACTTAACTGTAAGAGCATCTGGGTGTGAGAGTAAACACGCTCTGAA	4260
Db	4201	TGGATTTCAAACTTAACTGTAAGAGCATCTGGGTGTGAGAGTAAACACGCTCTGAA	4260
Oy	4261	ACTGCTCAGGGCTATCTCTCTCAGAGTGAATTTTAAACCTCAGCAGAGAGATACCATGC	4320
Db	4261	ACTGCTCAGGGCTATCTCTCTCAGAGTGAATTTTAAACCTCAGCAGAGAGATACCATGC	4320
Oy	4321	AACATTAACCTGTATAAGCTCCAGCAGAAATGGCGAATAGAACTGTGTTAAGAACAC	4380
Db	4321	AACATTAACCTGTATAAGCTCCAGCAGAAATGGCGAATAGAACTGTGTTAAGAACAC	4380
Oy	4381	ATGGAGGCAGCCTTTTAAACAGCTACCTTCCATATAGTACTCTTCTGCTTGGAG	4440
Db	4381	ATGGAGGCAGCCTTTTAAACAGCTACCTTCCATATAGTACTCTTCTGCTTGGAG	4440
Oy	4441	ACCTGCGAATTCAGAACAAACACATCAGAAAAGAGTATTTAACTTCACAGAAAAGTA	4500
Db	4441	ACCTGCGAATTCAGAACAAACACATCAGAAAAGAGTATTTAACTTCACAGAAAAGTA	4500
Oy	4501	GTGAATACCCATATAGCCAGATCCAGAAAGCCTTCTGCTGACAGTTTGAGGTGTCTG	4560
Db	4501	GTGAATACCCATATAGCCAGATCCAGAAAGCCTTCTGCTGACAGTTTGAGGTGTCTG	4560
Db	4501	GTGAATACCCATATAGCCAGATCCAGAAAGCCTTCTGCTGACAGTTTGAGGTGTCTG	4560
Oy	4561	CAGATAGTCTACAGATAAATAAAGAACACAGATGGAAGAGTATCCCTTCTAAT	4620
Db	4561	CAGATAGTCTACAGATAAATAAAGAACACAGATGGAAGAGTATCCCTTCTAAT	4620
Oy	4621	GCCCATCATTAAGATAGTGTGATACATGACAGTGTCTCTGAGAGTCTTCAAGATAGAA	4680
Db	4621	GCCCATCATTAAGATAGTGTGATACATGACAGTGTCTCTGAGAGTCTTCAAGATAGAA	4680
Oy	4681	ACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGACACAGCTGGAAG	4740
Db	4681	ACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGACACAGCTGGAAG	4740
Oy	4741	AGTCTGGGCACACAGATTGACGGAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA	4800
Db	4741	AGTCTGGGCACACAGATTGACGGAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA	4800
Oy	4801	CCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGATGACCTGAAATCTGATCCCTTG	4860
Db	4801	CCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGATGACCTGAAATCTGATCCCTTG	4860
Oy	4861	AAGACAGAGCCCGAGAGTCAAGTCTGTTGGCAATACCATCTTCAACCTCTGCATTTGA	4920
Db	4861	AAGACAGAGCCCGAGAGTCAAGTCTGTTGGCAATACCATCTTCAACCTCTGCATTTGA	4920
Oy	4921	AAGTCCCAATTTGAAGTTGCAAGTCTGCCAGAGTCCAGCTGCTGCTACTACTG	4980
Db	4921	AAGTCCCAATTTGAAGTCTGCCAGAGTCTGCCAGAGTCTGCTGCTACTACTG	4980
Oy	4981	ATACTCTGGGATTAATGCAATGGAAGAAAGTGTAGAGCAGGAGAGCAGAAATTCAG	5040
Db	4981	ATACTCTGGGATTAATGCAATGGAAGAAAGTGTAGAGCAGGAGAGCAGAAATTCAG	5040
Oy	5041	CTTCAACAGAAAGGCTCAACAAAGATTCATGATGCTGTCTGCTGACCCAGAG	5100
Db	5041	CTTCAACAGAAAGGCTCAACAAAGATTCATGATGCTGTGTGCTGACCCAGAG	5100
Oy	5101	AATTTATGCTGTGTGCAAGTTTGCAGAAACACACATCACTTAACTATCTAATTA	5160
Db	5101	AATTTATGCTGTGTGCAAGTTTGCAGAAACACACATCACTTAACTATCTAATTA	5160
Oy	5161	CTGAGAGACTACTCATGTTTGTATGAAAACAGATCTGAGTTGTGTGAACGGACAC	5220
Db	5161	CTGAGAGAGACTACTCATGTTTGTATGAAAACAGATCTGAGTTGTGTGAACGGACAC	5220
Oy	5221	TGAATATTTCTAGAAATTTGGGAGGAGAAATGGGTAGTATTTCTGGGTGACCC	5280
Db	5221	TGAATATTTCTAGAAATTTGGGAGGAGAAATGGGTAGTATTTCTGGGTGACCC	5280
Oy	5281	AGCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGG	5340
Db	5281	AGCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGG	5340
Oy	5341	TCAATGGAAGAAACACCAAGCTCCAAAGCAGCAAGAGATCCAGAGCAGAAAGATCT	5400
Db	5341	TCAATGGAAGAAACACCAAGCTCCAAAGCAGCAAGAGATCCAGAGCAGAAAGATCT	5400
Oy	5401	TCAGGGGGCTAGAAATCTGTGCTATGAGGCCCTTACCAACATGCCACAGATCACTGG	5460
Db	5401	TCAGGGGGCTAGAAATCTGTGCTATGAGGCCCTTACCAACATGCCACAGATCACTGG	5460
Oy	5461	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5520
Db	5461	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5520
Oy	5521	GCACAGGTGTCCACCAATTTGTGTTGTGACAGCAGATGCTGGAACAGAGACATGGCT	5580
Db	5521	GCACAGGTGTCCACCAATTTGTGTTGTGACAGCAGATGCTGGAACAGAGACATGGCT	5580
Oy	5581	TCCATGCAATTTGGGAGATGTTGTGAGGCACTGTGTGACCCGAGAGTGGGTGTTGACA	5640
Db	5581	TCCATGCAATTTGGGAGATGTTGTGAGGCACTGTGTGACCCGAGAGTGGGTGTTGACA	5640





Db 1081 GGGCTGGAAGTAMGAAACATGTAATGATAGCGGACTCCGACGACAGAAAAAGGTAG 1140  
QY 1141 ATCTGAATGCGATCCCTGTGTGAGAGAAAAAGATGGAATAGCAGAACTGCGATGCT 1200  
Db 1141 ATCTGAATGCGATCCCTGTGTGAGAGAAAAAGATGGAATAGCAGAACTGCGATGCT 1200  
QY 1201 CAGAGAACTCTAGAGATCTGGAAGATGTTCTTGATTAACACTAAATAGCAGATTCAGA 1260  
Db 1201 CAGAGAACTCTAGAGATCTGGAAGATGTTCTTGATTAACACTAAATAGCAGATTCAGA 1260  
QY 1261 AAGTATATGAGTGTGTTTCCAGAGGTAGTACACTGTTAGTTCGTGATGACTCACATGATG 1320  
Db 1261 AAGTATATGAGTGTGTTTCCAGAGGTAGTACACTGTTAGTTCGTGATGACTCACATGATG 1320  
QY 1321 GGGAGTGTGATCAATGCGAAATAGCTGATGATTGGAGCTCTTAATAGGAGATG 1380  
Db 1321 GGGAGTGTGATCAATGCGAAATAGCTGATGATTGGAGCTCTTAATAGGAGATG 1380  
QY 1381 AATATTCCTGTTCTTCAGAGAAATAGACTTACTGCGCAGTATCTCATGAGCTTTAA 1440  
Db 1381 AATATTCCTGTTCTTCAGAGAAATAGACTTACTGCGCAGTATCTCATGAGCTTTAA 1440  
QY 1441 TATGTAAAGTGAAGAGCTTCTCAATCAGATAGAGAGTAAATATTGAGACAAATAT 1500  
Db 1441 TATGTAAAGTGAAGAGCTTCTCAATCAGATAGAGAGTAAATATTGAGACAAATAT 1500  
QY 1501 TTGGGAAACCTATCGGAAGAGGACGCTCCCACTTAAGCCATGATGACCTGAAATC 1560  
Db 1501 TTGGGAAACCTATCGGAAGAGGACGCTCCCACTTAAGCCATGATGACCTGAAATC 1560  
QY 1561 TAAATATAGAGCATTTTGTACTGAGCCACAGATAAATACAGAGCGTCCCTCACAAATA 1620  
Db 1561 TAAATATAGAGCATTTTGTACTGAGCCACAGATAAATACAGAGCGTCCCTCACAAATA 1620  
QY 1621 AATTTAAAGCTTAAAGAGACCTTACATCAGGCTTCTCCTGAGAGATTTTATCAAGAAAG 1680  
Db 1621 AATTTAAAGCTTAAAGAGACCTTACATCAGGCTTCTCCTGAGAGATTTTATCAAGAAAG 1680  
QY 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATAAATCAGGAGACTAACCAAGGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAGACTCTGAAATGATAAATCAGGAGACTAACCAAGGAGC 1740  
QY 1741 AGATGCTCAAGTGAATATTACTAATAGTGTGTCATGAGATTAACCAAAAGGTGATT 1800  
Db 1741 AGATGCTCAAGTGAATATTACTAATAGTGTGTCATGAGATTAACCAAAAGGTGATT 1800  
QY 1801 CTAATTCAGATGAGAAATATCTTAACCCATAGAAATCAGTCAAGAAATCTGCTTTCA 1860  
Db 1801 CTAATTCAGATGAGAAATATCTTAACCCATAGAAATCAGTCAAGAAATCTGCTTTCA 1860  
QY 1861 AAAGCAAGCTGAACCTATTAAGCAGCTATTAAGCATATGGAATTCGAATTAATATCC 1920  
Db 1861 AAAGCAAGCTGAACCTATTAAGCAGCTATTAAGCATATGGAATTCGAATTAATATCC 1920  
QY 1921 ACAATTTCAAAAGCACTTAAAGAAATAGGCTGAGAGGAAAGTCTTTCACAGGCAATATTC 1980  
Db 1921 ACAATTTCAAAAGCACTTAAAGAAATAGGCTGAGAGGAAAGTCTTTCACAGGCAATATTC 1980  
QY 1981 ATGCGGCTGAAGTATGATAGTAAATGTAAGCCACCACTTAATTTGATGAAATTTGAAA 2040  
Db 1981 ATGCGGCTGAAGTATGATAGTAAATGTAAGCCACCACTTAATTTGATGAAATTTGAAA 2040  
QY 2041 TTGATAGTGTCTAGCAGTGAAGAGATTAAGAAAAAGTACAACTGAGCCAGATCA 2100  
Db 2041 TTGATAGTGTCTAGCAGTGAAGAGATTAAGAAAAAGTACAACTGAGCCAGATCA 2100  
QY 2101 GGCACAGCAGAAACCTTAACAATCATGAGGTAAGAACTGCAACTGAGGAGCAAGAGA 2160  
Db 2101 GGCACAGCAGAAACCTTAACAATCATGAGGTAAGAACTGCAACTGAGGAGCAAGAGA 2160  
QY 2161 GTRACAGGCAATGAAAGACAGCAAGTAAAGAGATGACATGACTTCTTCCAGAGCTGA 2220  
Db 2161 GTRACAGGCAATGAAAGACAGCAAGTAAAGAGATGACATGACTTCTTCCAGAGCTGA 2220  
QY 2221 AGTTAACAAATGACACCTGGTCTTTTCTTACTAGTGTCAATATACCAGTGAATTAAGAAAT 2280  
Db 2221 AGTTAACAAATGACACCTGGTCTTTTCTTACTAGTGTCAATATACCAGTGAATTAAGAAAT 2280  
QY 2281 TTGTCAATCTAGCCTTCCAGAGAGAGAAAAAGAGAAAGTAAAGAGTAAAGTGT 2340  
Db 2281 TTGTCAATCTAGCCTTCCAGAGAGAGAAAAAGAGAAAGTAAAGAGTAAAGTGT 2340  
QY 2341 CTAATATGCTGAAGAGCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAACCTG 2400  
Db 2341 CTAATATGCTGAAGAGCCCAAGATCTCATGTTAAGTGAAGAGGTTTTCGAACCTG 2400  
QY 2401 AAAGATCTGTAGAGAGTACAGTATTTCAATGCTAGCTGATGATTAATGAGCAGTCAAG 2460  
Db 2401 AAAGATCTGTAGAGAGTACAGTATTTCAATGCTAGCTGATGATTAATGAGCAGTCAAG 2460  
QY 2461 AAAGATCTGTAGAGAGTACAGTATTTCAATGCTAGCTGATGATTAATGAGCAGTCAAG 2520  
Db 2461 AAAGATCTGTAGAGAGTACAGTATTTCAATGCTAGCTGATGATTAATGAGCAGTCAAG 2520  
QY 2521 GTGTGAGTCAAGTGTGACAGCATTTTGAAGAACCCCAAGGAGTAAATTCATGTTTCCAAG 2580  
Db 2521 GTGTGAGTCAAGTGTGACAGCATTTTGAAGAACCCCAAGGAGTAAATTCATGTTTCCAAG 2580  
QY 2581 ATATATGAAATGACACAGAGGCTTAAAGTATCATTTGGAGATGAAATTAACACAGTTC 2640  
Db 2581 ATATATGAAATGACACAGAGGCTTAAAGTATCATTTGGAGATGAAATTAACACAGTTC 2640  
QY 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAACCTTATGATGCTGATTTGACAGATACAT 2700  
Db 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAACCTTATGATGCTGATTTGACAGATACAT 2700  
QY 2701 TCAAGGTTTCAAGGCGCATCATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
Db 2701 TCAAGGTTTCAAGGCGCATCATTTGCTGCTGTTTCAATCCAGAAATGAGAGAGG 2760  
QY 2761 AATGTGCAACATTCCTGCGCCACTGCGGCTTCTTAAAGAAACAAATCCAAAGTACATT 2820  
Db 2761 AATGTGCAACATTCCTGCGCCACTGCGGCTTCTTAAAGAAACAAATCCAAAGTACATT 2820  
QY 2821 TTGAATGTGAACAAAGAGAAATTCAGAGAAAGATGATGATTAATCAAGCTGTAC 2880  
Db 2821 TTGAATGTGAACAAAGAGAAATTCAGAGAAAGATGATGATTAATCAAGCTGTAC 2880  
QY 2881 AGACAGTTAATATCAGTCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
Db 2881 AGACAGTTAATATCAGTCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
QY 2941 ATGCAAAATGATATATCAAAAGAGGCTTAGTGTGCTATCAATCTCAGTTCAGAGCA 3000  
Db 2941 ATGCAAAATGATATATCAAAAGAGGCTTAGTGTGCTATCAATCTCAGTTCAGAGCA 3000  
QY 3001 ACGAACTGAGCTATTAATCTCAAAATGATGAGTCTTTTACAAACCCATATGCTATAC 3060  
Db 3001 ACGAACTGAGCTATTAATCTCAAAATGATGAGTCTTTTACAAACCCATATGCTATAC 3060  
QY 3061 CACCACCTTTTCCATCAAGTCAATTTGTTAAACTTAATGTAAGAAAAATCGCTAGAGG 3120  
Db 3061 CACCACCTTTTCCATCAAGTCAATTTGTTAAACTTAATGTAAGAAAAATCGCTAGAGG 3120  
QY 3121 AAAACTTTGAGAACTTCAATGTCACCTGAAAGAGAAATGGAATGGAATGGAATTCGCA 3180  
Db 3121 AAAACTTTGAGAACTTCAATGTCACCTGAAAGAGAAATGGAATGGAATGGAATTCGCA 3180  
QY 3181 GTACAGTGAAGCAATTAATGAGGTAATGAGAAATGTTTAAAGAGGAGCAGCT 3240  
Db 3181 GTACAGTGAAGCAATTAATGAGGTAATGAGAAATGTTTAAAGAGGAGCAGCT 3240  
QY 3241 CAAGCAATTAATTAAGAGTGTGCTGAGTAAATGAAAGTGGGCTCCATATTAATGAAA 3300  
Db 3241 CAAGCAATTAATTAAGAGTGTGCTGAGTAAATGAAAGTGGGCTCCATATTAATGAAA 3300

QY	3301	TAGGTTCCAGTGTGAAAACATTTCAAGCAGAACTAAGTAAACAAGAGGGCCAAATTTGA	3360
Db	3301	TAGGTTCCAGTGTGAAAACATTTCAAGCAGAACTAAGTAAACAAGAGGGCCAAATTTGA	3360
QY	3361	ATGCTATCTTAAATTAGGGGTTTTGCAACCTGAGGGTATATAACAAAGTCTCTGGAA	3420
Db	3361	ATGCTATCTTAAATTAGGGGTTTTGCAACCTGAGGGTATATAACAAAGTCTCTGGAA	3420
QY	3421	GTAATTGTAACATCTCGAATAATAAAAAGCAAGATATGAAGAGTAACTGACTGTA	3480
Db	3421	GTAATTGTAACATCTCGAATAATAAAAAGCAAGATATGAAGAGTAACTGACTGTA	3480
QY	3481	ATPACAGATTTCTTCACATATCTGATTTTCAGATTAACCTTAGAAGCCTTAGGGAACTACTC	3540
Db	3481	ATPACAGATTTCTTCACATATCTGATTTTCAGATTAACCTTAGAAGCCTTAGGGAACTACTC	3540
QY	3541	ATGCAATCTCAGAGTTGTTCTGAGACACCTGATGACCTGTATGATGATGCTAAATTAAGG	3600
Db	3541	ATGCAATCTCAGAGTTGTTCTGAGACACCTGATGACCTGTATGATGATGCTAAATTAAGG	3600
QY	3601	AAGATACTAGTTTTCGTGAAAAATACATTAAGAAAGATTCTGCTGTTTTTACGAAAACGC	3660
Db	3601	AAGATACTAGTTTTCGTGAAAAATACATTAAGAAAGATTCTGCTGTTTTTACGAAAACGC	3660
QY	3661	TTCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTACCACATACACATTTGGCTCAGG	3720
Db	3661	TTCCAGAGAGGAGAGCTTAGCAGAGAGTCTAGCCCTTACCACATACACATTTGGCTCAGG	3720
QY	3721	GTTACCGAAGAGGGGGCCAGAAATTAAGAGTCTCAGAAAGAACTTATCTAGTAGAGATG	3780
Db	3721	GTTACCGAAGAGGGGGCCAGAAATTAAGAGTCTCAGAAAGAACTTATCTAGTAGAGATG	3780
QY	3781	AAGAGCTTCCCTGCTCCACACACTTGTATTGTGTAAGTAAACAAATTAACCTTCTCAGT	3840
Db	3781	AAGAGCTTCCCTGCTCCACACACTTGTATTGTGTAAGTAAACAAATTAACCTTCTCAGT	3840
QY	3841	CTACTAGGCAATAGACCGGTCTCTACCGAGTGTCTGTAGACACAGAGAGATTTAT	3900
Db	3841	CTACTAGGCAATAGACCGGTCTCTACCGAGTGTCTGTAGACACAGAGAGATTTAT	3900
QY	3901	TATCATTTGAAGATAGCTTAATATAGACAGTAACCCAGGTAAATTTGSCAAAGGCAATCTC	3960
Db	3901	TATCATTTGAAGATAGCTTAATATAGACAGTAACCCAGGTAAATTTGSCAAAGGCAATCTC	3960
QY	3961	AGGAACATCACCTTAGTGAAGGAAACAAATTTCTGCTAGCTTTTCTTCCACAGTGA	4020
Db	3961	AGGAACATCACCTTAGTGAAGGAAACAAATTTCTGCTAGCTTTTCTTCCACAGTGA	4020
QY	4021	GTGAATTGGAAGACTTGACTGCAAAATACAAACCCAGAGATCTTCTGATTTGGTCTT	4080
Db	4021	GTGAATTGGAAGACTTGACTGCAAAATACAAACCCAGAGATCTTCTGATTTGGTCTT	4080
QY	4081	CCAAACAAATGAGGCATCTAGTCTAAAGCCAGGGAGTTGGTCTAGTGACAAAGAAATTTGG	4140
Db	4081	CCAAACAAATGAGGCATCTAGTCTAAAGCCAGGGAGTTGGTCTAGTGACAAAGAAATTTGG	4140
QY	4141	TTTTCAAGATGATGAAGAAAGAGAACCGGGCTTGAAGAAATATATCAAGAACAGCAAGCA	4200
Db	4141	TTTTCAAGATGATGAAGAAAGAGAACCGGGCTTGAAGAAATATATCAAGAACAGCAAGCA	4200
QY	4201	TGGATTCAAACTTAGTGTAAGCAGACATCTGGGTGTAGAGTGAACAAGCGTCTCTGAG	4260
Db	4201	TGGATTCAAACTTAGTGTAAGCAGACATCTGGGTGTAGAGTGAACAAGCGTCTCTGAG	4260
QY	4261	ACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAACCACTCAGACAGAGGATACCATGC	4320
Db	4261	ACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAACCACTCAGACAGAGGATACCATGC	4320
QY	4321	AACCTTAACCTGATTAAGCTCCAGAGAAATGCGTGAACCTTAGAGCTGTGTAGAACAGC	4380
Db	4321	AACCTTAACCTGATTAAGCTCCAGAGAAATGCGTGAACCTTAGAGCTGTGTAGAACAGC	4380
QY	4381	ATGGAGCCAGCTTCTTAACAGCTTACCTTCCATATAGTACTTCTTCCCTTGAGG	4440

Db	4381	ATGGGAGCCAGCCTTCTTAACACCTACCCCTTCATCATATAAGTACCTCTCTGCCCCTTGAGG	4444
OY	4441	ACCTGGAAATCCAGAAACAAGACATCGAAGAAAGACAGTATTAACTTACACGAAACAGA	4500
Db	4441	ACCTGCGAAATCCAGAACACACATCAGAAAAAGCAATTAACCTTCACGAAAAAGTA	4500
OY	4501	GTAATATACCTTAAAGCCAGAAATCCAGAAAGCCTTCTCTGACAAAGTTTGAGTGTCTG	4560
Db	4501	GTAATATACCTTAAAGCCAGAAATCCAGAAAGCCTTCTCTGACAAAGTTTGAGTGTCTG	4560
OY	4561	CAGATATGTTCTACCACTATAAATAAAGAACAGAGGTGAAAGTCAATCCCTTTAAAT	4620
Db	4561	CAGATATGTTCTACCACTATAAATAAAGAACAGAGGTGAAAGTCAATCCCTTTAAAT	4620
OY	4621	GGCCATCATTTACATATAGTGTGTCATATGACAGATTTGGCTCTGGAGTCTTCAGATTAGA	4680
Db	4621	GGCCATCATTTACATATAGTGTGTCATATGACAGATTTGGCTCTGGAGTCTTCAGATTAGA	4680
OY	4681	ACTACCCATCTCAAGAGAGCACTTAAGTTTGATGTGGAGAGCAACAGCTGGAAG	4740
Db	4681	ACTACCCATCTCAAGAGAGCACTTAAGTTTGATGTGGAGAGCAACAGCTGGAAG	4740
OY	4741	AATCTGGGCGACACATTTGACGGAACATCTTACTTGCCAAAGCAAGATCTAGAGGA	4800
Db	4741	AATCTGGGCGACACATTTGACGGAACATCTTACTTGCCAAAGCAAGATCTAGAGGA	4800
OY	4801	CCCCCTTACCTGGAATCTGGAATCACCCTCTCTCTGATGACCTGAATCTGATCCTCTG	4860
Db	4801	CCCCCTTACCTGGAATCTGGAATCACCCTCTCTCTGATGACCTGAATCTGATCCTCTG	4860
OY	4861	AAGACAGAGCCCCACAGTACGCTGCTGTTGGCAATACCAATCTTCAACCTCTGCATGGA	4920
Db	4861	AAGACAGAGCCCCACAGTACGCTGCTGTTGGCAATACCAATCTTCAACCTCTGCATGGA	4920
OY	4921	AAGTTCCCAATTGAAAGTTGCGAATCTGCCACAGGTCACGCTCTCTCACTACTCTG	4980
Db	4921	AAGTTCCCAATTGAAAGTTGCGAATCTGCCACAGGTCACGCTCTCTCACTACTCTG	4980
OY	4981	ATAGTGTGGGTATATGCAATGGAAGAAAGTGTGAGCAGGGAAGGACAGAAATTGACAG	5040
Db	4981	ATAGTGTGGGTATATGCAATGGAAGAAAGTGTGAGCAGGGAAGGACAGAAATTGACAG	5040
OY	5041	CTTCAACAGAAAGGGTCAACAACAAATGTCATGTTGCTGCTGACCCGACAGAG	5100
Db	5041	CTTCAACAGAAAGGGTCAACAACAAATGTCATGTTGCTGCTGACCCGACAGAG	5100
OY	5101	AATTTATGCTGTGTCACAGTTTGCAGAAACACACATCATCTTAATCTTAATTA	5160
Db	5101	AATTTATGCTGTGTCACAGTTTGCAGAAACACACATCATCTTAATCTTAATTA	5160
OY	5161	CGAAGAGACCTACATGTTGTTATGAAGAAACGATGTGAGTTGTGTGAACGGACAC	5220
Db	5161	CGAAGAGACCTACATGTTGTTATGAAGAAACGATGTGAGTTGTGTGAACGGACAC	5220
OY	5221	TGAATATATTTCTAGAGAAATGGGGAGGAAATGGGTAGTATGATTTCTGGGTGACCC	5280
Db	5221	TGAATATATTTCTAGAGAAATGGGGAGGAAATGGGTAGTATGATTTCTGGGTGACCC	5280
OY	5281	AGTCTATTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGTGG	5340
Db	5281	AGTCTATTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGTGG	5340
OY	5341	TCAATGGAAGAAACCCACCAAGTCCAAAGCGAAGAAATCCAGACAGAAAGATCT	5400
Db	5341	TCAATGGAAGAAACCCACCAAGTCCAAAGCGAAGAAATCCAGACAGAAAGATCT	5400
OY	5401	TCAGGGGCTAGAAATCTGTTGCTATGGGCCCTTACCAACATGCGCCACATCAATCAGG	5460
Db	5401	TCAGGGGCTAGAAATCTGTTGCTATGGGCCCTTACCAACATGCGCCACATCAATCAGG	5460
OY	5461	ATAGATGTGACAGCTGTGTGTGCTTGTGTGTAAGAGCTTTCATCATTCACCTCTG	5520
Db	5461	ATAGATGTGACAGCTGTGTGTGCTTGTGTGTAAGAGCTTTCATCATTCACCTCTG	5520

```

Db      5461 AATGATGATGACAGCTGTGTGCTCTTCTGTGTGAAGAGCTTTCATTCACCTTGG 5520
QY      5521 GCACAGGTGTCACCAATGTGTGTGTGCGACGACAGATCCTGACAGAGAGACATGGCT 5580
Db      5521 GCACAGGTGTCACCAATGTGTGTGTGCGACGACAGATCCTGACAGAGAGATGGCT 5580
QY      5581 TCCATGCAATTTGGGACAGATGTGTGAGGACACCTGTGTGACCCGAGAGTGGTGTGACA 5640
Db      5581 TCCATGCAATTTGGGACAGATGTGTGAGGACACCTGTGTGACCCGAGAGTGGTGTGACA 5640
QY      5641 GTGTAGACCTCTTACCACTGCGACAGAGCTGACACCTTACCTGATACCCAGATCCGCCACA 5700
Db      5641 GTGTAGACCTCTTACCACTGCGACAGAGCTGACACCTTACCTGATACCCAGATCCGCCACA 5700
QY      5701 GCCACTACTGA 5711
Db      5701 GCCACTACTGA.5711

```

## RESULT 7

```

US-09-982-828-1
; Sequence 1, Application US/09982828
; Publication No. US2003022184A1

```

## GENERAL INFORMATION:

```

APPLICANT: Murphy, Patricia D.
           Allen, Antonette C.
           Alvarez, Christopher P.
           Critz, Brenda S.
           Olson, Sheri J.
           Thurber, Denise
           Zeng, Bin

```

```

TITLE OF INVENTION: Coding Sequences of the Human
                    BRCA1 Gene

```

```

NUMBER OF SEQUENCES: 72

```

## CORRESPONDENCE ADDRESSES:

```

ADDRESSEE: Morgan Lewis & Bockius LLP
STREET: 1111 Pennsylvania Avenue N. W.
CITY: Washington
STATE: DC

```

```

COUNTRY: USA

```

```

ZIP: 20004

```

## COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/09/982.828

```

```

FILING DATE: 22-Oct-2001

```

```

CLASSIFICATION: <unknown>

```

## PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 09/074,453

```

```

FILING DATE: 1998-05-06

```

```

APPLICATION NUMBER: US 08/798,691

```

```

FILING DATE: 1997-02-12

```

```

APPLICATION NUMBER: US 08/598,591

```

```

FILING DATE: 1996-02-12

```

## ATTORNEY/AGENT INFORMATION:

```

NAME: Michael S. Tuscan

```

```

REGISTRATION NUMBER: 43,210

```

```

REFERENCE/DOCKET NUMBER: 44921-5053-01-US

```

## TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 202-739-3000

```

```

TELEFAX: 202-739-3001

```

## INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 5711 base pairs

```

```

TYPE: nucleic acid

```

```

STRANDEDNESS: not relevant

```

```

TOPOLOGY: linear

```

```

MOLECULE TYPE: cDNA

```

```

ORGANISM: Homo sapiens

```

```

; STRAIN: BRCA1 (om11)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-982-828-1

```

```

Query Match      99.9% Score 5703; DB 9; Length 5711:
Best Local Similarity 99.9% Pred. No. 0;
Matches 5706; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 AGCTGCTGAGACTTCCCTGGACCCCGACAGGCTGTGGGTTTCTCAGATAACCTGGCC 60
Db      1 AGCTGCTGAGACTTCCCTGGACCCCGACAGGCTGTGGGTTTCTCAGATAACCTGGCC 60
QY      61 CCTGCGCTCAGAGAGGCTTACCCCTGCTCTGTGGGTAAGTTATTGGAACAGAAAGAA 120
Db      61 CCTGCGCTCAGAGAGGCTTACCCCTGCTCTGTGGGTAAGTTATTGGAACAGAAAGAA 120
QY      121 TGSATTAATCGCTCTCGGGTTGAAGATGACAAATGCAATTAATGCTATGACAGAAA 180
Db      121 TGSATTAATCGCTCTCGGGTTGAAGATGACAAATGCAATTAATGCTATGACAGAAA 180
QY      181 TCTTAGAGTGCCCATCTGCTGTGAGTTGATCAAGAGAACCTGTCTCCAGAAAGTGACC 240
Db      181 TCTTAGAGTGCCCATCTGCTGTGAGTTGATCAAGAGAACCTGTCTCCAGAAAGTGACC 240
QY      241 ACATATTTTGAATTTTGTGATGCTGCAACTTCTCAACCAAGAAAGAGGCTTCAAGT 300
Db      241 ACATATTTTGAATTTTGTGATGCTGCAACTTCTCAACCAAGAAAGAGGCTTCAAGT 300
QY      301 GTCCCTTATGTAAGATGATATTAACCAAGAGGCTTCAAGAAAGATGAGATTAGTC 360
Db      301 GTCCCTTATGTAAGATGATATTAACCAAGAGGCTTCAAGAAAGATGAGATTAGTC 360
QY      361 AACTTGTGAAGAGCTATTAATAATATTTGTCTTTGACGTTGACAGGCTTTGAGT 420
Db      361 AACTTGTGAAGAGCTATTAATAATATTTGTCTTTGACGTTGACAGGCTTTGAGT 420
QY      421 ATGCAAAACAGCTATATTTTGCAGAAAAGAAAATACCTCTCTGAACATCTAAAGATG 480
Db      421 ATGCAAAACAGCTATATTTTGCAGAAAAGAAAATACCTCTCTGAACATCTAAAGATG 480
QY      481 AAGTTCTATCACCAGATGATGGGCTACGAAACCGTGCAAAACACTTCTACAGAGT 540
Db      481 AAGTTCTATCACCAGATGATGGGCTACGAAACCGTGCAAAACACTTCTACAGAGT 540
QY      541 AACCCGAAAATCCTCTCTGACAGAAAACAGTCTCAAGTCTTAACCTTGGAA 600
Db      541 AACCCGAAAATCCTCTCTGACAGAAAACAGTCTCAAGTCTTAACCTTGGAA 600
QY      601 CTGTGAGAACTGTGAGAGCAAAAGCAGGATACAACTTCAAAAAGAGCTGTCTACATTG 660
Db      601 CTGTGAGAACTGTGAGAGCAAAAGCAGGATACAACTTCAAAAAGAGCTGTCTACATTG 660
QY      661 AATTGGATCTGATTTCTTGAAGATACGCTTAATAGGCAACTTAATTTCAAGTGTGGAG 720
Db      661 AATTGGATCTGATTTCTTGAAGATACGCTTAATAGGCAACTTAATTTCAAGTGTGGAG 720
QY      721 ATCAGAAATTTGTAACAATCACCCCTCAAGAAACAGGATGAATCAATTTGGATTCTG 780
Db      721 ATCAGAAATTTGTAACAATCACCCCTCAAGAAACAGGATGAATCAATTTGGATTCTG 780
QY      781 CAAAAAAGGCTGCTTGTGAATTTTGTGAGAGGATGTAACAAATACCTGATCAAC 840
Db      781 CAAAAAAGGCTGCTTGTGAATTTTGTGAGAGGATGTAACAAATACCTGATCAAC 840
QY      841 CCAATATATGATTTGAACACACGAGAGGCTCAGTACAGAGGCTCAGAGAAAGT 900
Db      841 CCAATATATGATTTGAACACACGAGAGGCTCAGTACAGAGGCTCAGAGAAAGT 900
QY      901 ATCAGGTAAGTTGTTCAAACTGCAATGTGAGGCAATGTGCAAAATACATGCA 960
Db      901 ATCAGGTAAGTTGTTCAAACTGCAATGTGAGGCAATGTGCAAAATACATGCA 960

```

Db 901 ATCAGGTAGTTCGTTTCAAACTTGATGAGCCATGTGGCAAAATACATCGCA 960  
Qy 961 GCTCATTCAGCATGAGAACAGCAGCTTATTAATCTACATAAAGACGAAATGATAGAAA 1020  
Db 961 GCTCATTCAGCATGAGAACAGCAGCTTATTAATCTACATAAAGACGAAATGATAGAAA 1020  
Qy 1021 AGGCTGAATTCGTATATAAAGCAAAACAGCCTGGCTTACGAAGAGCCACATACAGAT 1080  
Db 1021 AGGCTGAATTCGTATATAAAGCAAAACAGCCTGGCTTACGAAGAGCCACATACAGAT 1080  
Qy 1081 GGGCTGGAAGTAGAAGAAACATGTATATAGCGGAGCTCCAGCAGAGAAAAAGTAG 1140  
Db 1081 GGGCTGGAAGTAGAAGAAACATGTATATAGCGGAGCTCCAGCAGAGAAAAAGTAG 1140  
Qy 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGAGAGAAATGCCATGCT 1200  
Db 1141 ATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGAGAGAAATGCCATGCT 1200  
Qy 1201 CAGAGAAATCCTAGAGATCTGAGATGTTCTTGATTAACACTAAATAGCAGATTCAGA 1260  
Db 1201 CAGAGAAATCCTAGAGATCTGAGATGTTCTTGATTAACACTAAATAGCAGATTCAGA 1260  
Qy 1261 AAGTTAATGAGTGGTTTCCAGAGAGTGAATGTAAGTTAGTTCTGATCAGATGATG 1320  
Db 1261 AAGTTAATGAGTGGTTTCCAGAGAGTGAATGTAAGTTAGTTCTGATCAGATGATG 1320  
Qy 1321 GGGAGTCTGAATCCAAATGCCAAAGTAGATGATGATGAGAGTCTTAATAGAGTAGTG 1380  
Db 1321 GGGAGTCTGAATCCAAATGCCAAAGTAGATGATGATGAGAGTCTTAATAGAGTAGTG 1380  
Qy 1381 AATATTCGTGTTCTTCAGAGAAATAGACTTACTGCGCAGTATCCATGAGGCTTTAA 1440  
Db 1381 AATATTCGTGTTCTTCAGAGAAATAGACTTACTGCGCAGTATCCATGAGGCTTTAA 1440  
Qy 1441 TATGTAAAAAGTGAAGAGTTCACATCAATAGTAGAGATTAATTAAGCAAAATAT 1500  
Db 1441 TATGTAAAAAGTGAAGAGTTCACATCAATAGTAGAGATTAATTAAGCAAAATAT 1500  
Qy 1501 TTGGGAAAACCTATCGAAGAGGACGCTCCCAACTTAAGCCATGTAATCTGAAAATC 1560  
Db 1501 TTGGGAAAACCTATCGAAGAGGACGCTCCCAACTTAAGCCATGTAATCTGAAAATC 1560  
Qy 1561 TAAATTAAGAGACATTTGTACTGAGCCAGATTAATCAAGAGGCTCCCTCACAAATA 1620  
Db 1561 TAAATTAAGAGACATTTGTACTGAGCCAGATTAATCAAGAGGCTCCCTCACAAATA 1620  
Qy 1621 AATTAACCGTAAAAAGAGACCTTACATGAGGCTTCACTGAGGATTTTCAAGAAAG 1680  
Db 1621 AATTAACCGTAAAAAGAGACCTTACATGAGGCTTCACTGAGGATTTTCAAGAAAG 1680  
Qy 1681 CAGATTTGGCAGTTCAAAAGAGCTCCTGAATGATTAATCAGGGAATCAACCAAGGAGC 1740  
Db 1681 CAGATTTGGCAGTTCAAAAGAGCTCCTGAATGATTAATCAGGGAATCAACCAAGGAGC 1740  
Qy 1741 AGAATGGTCAAGTATGATTAATTAATGATGATGATGATGATGATGATGATGATGAT 1800  
Db 1741 AGAATGGTCAAGTATGATTAATTAATGATGATGATGATGATGATGATGATGATGAT 1800  
Qy 1801 CTATTCAGATGAGAAAAATCCTTAACCAATGAATCACTGAAAAAATCTGCTTTCA 1860  
Db 1801 CTATTCAGATGAGAAAAATCCTTAACCAATGAATCACTGAAAAAATCTGCTTTCA 1860  
Qy 1861 AAAGGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAATGGAATTAATACC 1920  
Db 1861 AAAGGAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAATGGAATTAATACC 1920  
Qy 1921 ACAATTCAAAAGCACTTAAGAAATAGCTGAGAGAGAGCTTCTTACCAAGCATATTC 1980  
Db 1921 ACAATTCAAAAGCACTTAAGAAATAGCTGAGAGAGAGCTTCTTACCAAGCATATTC 1980  
Qy 1981 ATGGCTTGAACTAGTACGATGAGAAATCTTAAGCCCACTAATTTGTAATTTGCCAA 2040  
Db 1981 ATGGCTTGAACTAGTACGATGAGAAATCTTAAGCCCACTAATTTGTAATTTGCCAA 2040

Qy 2041 TTGATAGTGTCTTCAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCCAGTCA 2100  
Db 2041 TTGATAGTGTCTTCAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCCAGTCA 2100  
Qy 2101 GGCACAGCAGAAACCTTACACTCATGGAAGGTAAAGAACTGCAACTGGAGCCAGAAAGA 2160  
Db 2101 GGCACAGCAGAAACCTTACACTCATGGAAGGTAAAGAACTGCAACTGGAGCCAGAAAGA 2160  
Qy 2161 GTACACACCCTAATGACAGACAGTAAAGACATGACAGTGAATCTTTCCAGAGCTGA 2220  
Db 2161 GTACACACCCTAATGACAGACAGTAAAGACATGACAGTGAATCTTTCCAGAGCTGA 2220  
Qy 2221 AGTTAAACAAATGACACTGCTGTTCTTTACTAAGTGTCAATACCGTAGTAAATAGAT 2280  
Db 2221 AGTTAAACAAATGACACTGCTGTTCTTTACTAAGTGTCAATACCGTAGTAAATAGAT 2280  
Qy 2281 TTGTCAATCTAGACCTTCAGAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2340  
Db 2281 TTGTCAATCTAGACCTTCAGAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2340  
Qy 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGGCAACTG 2400  
Db 2341 CTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGGCAACTG 2400  
Qy 2401 AAAGATCTGTAGAGTAGCAGTATTTGATGTTACCTGATGATGATGATGATGATGATG 2460  
Db 2401 AAAGATCTGTAGAGTAGCAGTATTTGATGTTACCTGATGATGATGATGATGATGATG 2460  
Qy 2461 AAAGATCTGTAGAGTAGCAGTATTTGATGTTACCTGATGATGATGATGATGATGATG 2520  
Db 2461 AAAGATCTGTAGAGTAGCAGTATTTGATGTTACCTGATGATGATGATGATGATGATG 2520  
Qy 2521 GTGTGAGTCAAGTGTGACACATTTGAAAACCCCAAGGACTAATTCATGATGTTCCAAAG 2580  
Db 2521 GTGTGAGTCAAGTGTGACACATTTGAAAACCCCAAGGACTAATTCATGATGTTCCAAAG 2580  
Qy 2581 ATTAATGAATGACACAGAGGCTTAAGTATGATGATGATGATGATGATGATGATGATG 2640  
Db 2581 ATTAATGAATGACACAGAGGCTTAAGTATGATGATGATGATGATGATGATGATGATG 2640  
Qy 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGATGATGATGATGATGATGATGATGAT 2700  
Db 2641 GGGAAACAGCATAGAAATGGAAGAAAGTGAATGATGATGATGATGATGATGATGATGAT 2700  
Qy 2701 TCAAGGTTTCAAGGCGCAGTATTTGCTGTTTCAAAATCCAGAAATGAGAAAGAG 2760  
Db 2701 TCAAGGTTTCAAGGCGCAGTATTTGCTGTTTCAAAATCCAGAAATGAGAAAGAG 2760  
Qy 2761 AATGTGCAACATTTCTCTGCCACCTGCGCTTAAAGAAACAAAGTCCAAAGCTCAT 2820  
Db 2761 AATGTGCAACATTTCTCTGCCACCTGCGCTTAAAGAAACAAAGTCCAAAGCTCAT 2820  
Qy 2821 TTGAATGTGAACCAAAAGAGAAATCAAGAAAGATGATGATGATGATGATGATGATGAT 2880  
Db 2821 TTGAATGTGAACCAAAAGAGAAATCAAGAAAGATGATGATGATGATGATGATGATGAT 2880  
Qy 2881 AGACAGTTAATATCAGCTGAGGCTTCTGTGTTGTCAGAAAGATAGCCAGTTGATA 2940  
Db 2881 AGACAGTTAATATCAGCTGAGGCTTCTGTGTTGTCAGAAAGATAGCCAGTTGATA 2940  
Qy 2941 ATGGCAATGTAGATATCAAGAGAGGCTGATGTTTGTCTATCAATCTAGTTCAGAGGCA 3000  
Db 2941 ATGGCAATGTAGATATCAAGAGAGGCTGATGTTTGTCTATCAATCTAGTTCAGAGGCA 3000  
Qy 3001 AGCAAACTGAGACTATTAATCAATTAAGTGAATGCAATTAAGCAAAACCATATGATATC 3060  
Db 3001 AGCAAACTGAGACTATTAATCAATTAAGTGAATGCAATTAAGCAAAACCATATGATATC 3060  
Qy 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAGG 3120  
Db 3061 CACCACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAGG 3120

QY 3121 AAAACCTTGGAGCAATTCACATGTCACCTGAAAGAGAAATGGGAAATGAGAAATCCAA 3180  
Db 3121 AAAACCTTGGAGCAATTCACATGTCACCTGAAAGAGAAATGGGAAATGAGAAATCCAA 3180  
QY 3181 GTACAGGAGCACAATTAAGCCGTAAATACATTAGAGAAATGTTTAAAGAGCCACT 3240  
Db 3181 GTACAGGAGCACAATTAAGCCGTAAATACATTAGAGAAATGTTTAAAGAGCCACT 3240  
QY 3241 CAACCAATTAATGAAGTACCTGCTCAATGAAGTGGGCTCCAGTATTAATGAA 3300  
Db 3241 CAACCAATTAATGAAGTACCTGCTCAATGAAGTGGGCTCCAGTATTAATGAA 3300  
QY 3301 TAGGTTCCAGTATGAGAAACATTCAGAGAACTAGTAGAAGAGAGGCCAAATTA 3360  
Db 3301 TAGGTTCCAGTATGAGAAACATTCAGAGAACTAGTAGAAGAGAGGCCAAATTA 3360  
QY 3361 ATGCTATGCTAGATTTAGGGGTTTGGACCTGAGGTCTATTAACAAAGTCTTCTGAA 3420  
Db 3361 ATGCTATGCTAGATTTAGGGGTTTGGACCTGAGGTCTATTAACAAAGTCTTCTGAA 3420  
QY 3421 GTAATTTAGCAATCTGAAATTAAGAAAGCAAGATATGAGAGTATTCAGACTTTA 3480  
Db 3421 GTAATTTAGCAATCTGAAATTAAGAAAGCAAGATATGAGAGTATTCAGACTTTA 3480  
QY 3481 ATACAGATTTCTCTCCATATCTGATTCAGATTAACACTAGAACAGCTATGSGAAGTATC 3540  
Db 3481 ATACAGATTTCTCTCCATATCTGATTCAGATTAACACTAGAACAGCTATGSGAAGTATC 3540  
QY 3541 ATGCAATCTAGGTTGTTCTGAGACCTGATGACCTGTTAGATGATGTAATTAAGG 3600  
Db 3541 ATGCAATCTAGGTTGTTCTGAGACCTGATGACCTGTTAGATGATGTAATTAAGG 3600  
QY 3601 AAGATACTAGTTTGTGAAATAGACATTAAGAAAGTTTGTCTGTTTAAAGCAAAAGCG 3660  
Db 3601 AAGATACTAGTTTGTGAAATAGACATTAAGAAAGTTTGTCTGTTTAAAGCAAAAGCG 3660  
QY 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTACGCCCTTTCACCATATACATTTGGCTAGG 3720  
Db 3661 TCCAGAAAGAGAGCTTAGCAGAGAGTCTACGCCCTTTCACCATATACATTTGGCTAGG 3720  
QY 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAGACATTAATCTGTAGAGTG 3780  
Db 3721 GTTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAGACATTAATCTGTAGAGTG 3780  
QY 3781 AAGAGCTTCCCTGCTTCCAAACACTTGTATTTGGTAAAGTAACAAATATACCTTCTCAGT 3840  
Db 3781 AAGAGCTTCCCTGCTTCCAAACACTTGTATTTGGTAAAGTAACAAATATACCTTCTCAGT 3840  
QY 3841 CTACTAGGCAATAGCAGCGTGTCTACGAGTGTCTGTAGAACACAGAGAGAAATTTAT 3900  
Db 3841 CTACTAGGCAATAGCAGCGTGTCTACGAGTGTCTGTAGAACACAGAGAGAAATTTAT 3900  
QY 3901 TATCATTTGAGAAATAGCTTAATAGTACGTACAGTAACAGGTAAATTTGGCAAGGCAATTC 3960  
Db 3901 TATCATTTGAGAAATAGCTTAATAGTACGTACAGTAACAGGTAAATTTGGCAAGGCAATTC 3960  
QY 3961 AGGAACATCACTTGTAGAGAGAAACAAATGTTCTGTAGCTTTTCTTCAACAGTGCA 4020  
Db 3961 AGGAACATCACTTGTAGAGAGAAACAAATGTTCTGTAGCTTTTCTTCAACAGTGCA 4020  
QY 4021 GTGAATTTGGAAGACTTACTGCAATATCAAAACCCAGAGATCTTCTTCAACAGTGCA 4080  
Db 4021 GTGAATTTGGAAGACTTACTGCAATATCAAAACCCAGAGATCTTCTTCAACAGTGCA 4080  
QY 4081 CCAAAACAATGAGGCAATGCTGAAGCCAGGAGTGGTCTGAGTGAAGCAAGAAATGG 4140  
Db 4081 CCAAAACAATGAGGCAATGCTGAAGCCAGGAGTGGTCTGAGTGAAGCAAGAAATGG 4140  
QY 4141 TTTTCAGATGATGAAGAAAGAGAAAGGCTTGGAAAGAAATTAATCAAGAGAGCAAGCA 4200  
Db 4141 TTTTCAGATGATGAAGAAAGAGAAAGGCTTGGAAAGAAATTAATCAAGAGAGCAAGCA 4200  
QY 4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAGCAAGCGCTCTGAGG 4260

Db 4201 TGGATTTCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAGCAAGCGCTCTGAGG 4260  
QY 4261 ACTGCTCAGGCGTATTCCTCTAGAGTGAATTTAAACCACTCAGCAGAGGATACCATG 4320  
Db 4261 ACTGCTCAGGCGTATTCCTCTAGAGTGAATTTAAACCACTCAGCAGAGGATACCATG 4320  
QY 4321 AACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGACTAGAAAGCTGTTTGAACAGC 4380  
Db 4321 AACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGACTAGAAAGCTGTTTGAACAGC 4380  
QY 4381 ATGGAGCCAGCTTCTTAACAGCTAACCTTCCATCATTAAGTGAAGTCTTCTGAGG 4440  
Db 4381 ATGGAGCCAGCTTCTTAACAGCTAACCTTCCATCATTAAGTGAAGTCTTCTGAGG 4440  
QY 4441 ACCTGCGAAATCCGAGCAAAAGCAATCAGAAAAAGAGATTAATTAATCAACAAAGTA 4500  
Db 4441 ACCTGCGAAATCCGAGCAAAAGCAATCAGAAAAAGAGATTAATTAATCAACAAAGTA 4500  
QY 4501 GTGAATACCCATTAAGCCAGATCCAGAAAGGCTTCTGCTGCAAGTGTGAGGTCTG 4560  
Db 4501 GTGAATACCCATTAAGCCAGATCCAGAAAGGCTTCTGCTGCAAGTGTGAGGTCTG 4560  
QY 4561 CAGATAGTCTTACAGTAAATTAAGAACCCAGAGTGAAGGTCATCCCTTTCTAAT 4620  
Db 4561 CAGATAGTCTTACAGTAAATTAAGAACCCAGAGTGAAGGTCATCCCTTTCTAAT 4620  
QY 4621 GCCCATCATTTAGATGATGAGTGTGATGACAGATGCTTCTGAGAGCTTCAAAATGAA 4680  
Db 4621 GCCCATCATTTAGATGATGAGTGTGATGACAGATGCTTCTGAGAGCTTCAAAATGAA 4680  
QY 4681 ACTACCCATCTCAAGAGAGAGCTTAATAGGTTGTTGATGTGAGAGCAACAGCTGAGG 4740  
Db 4681 ACTACCCATCTCAAGAGAGAGCTTAATAGGTTGTTGATGTGAGAGCAACAGCTGAGG 4740  
QY 4741 AGTCTGGGCCACACGATTTGACGAAACATCTTACTTCCCAAGGCAAGATCTAGAGGAA 4800  
Db 4741 AGTCTGGGCCACACGATTTGACGAAACATCTTACTTCCCAAGGCAAGATCTAGAGGAA 4800  
QY 4801 CCCCTTACCTGGAATCTGGAATCAGGCTCTTCTGTAAGGACCTGGAATGATCTTCTG 4860  
Db 4801 CCCCTTACCTGGAATCTGGAATCAGGCTCTTCTGTAAGGACCTGGAATGATCTTCTG 4860  
QY 4861 AAGACAGAGCCCAAGATCAGCTGCTGTGGAACAATACCATTTCACTGCAATGA 4920  
Db 4861 AAGACAGAGCCCAAGATCAGCTGCTGTGGAACAATACCATTTCACTGCAATGA 4920  
QY 4921 AAGTCCCAATTAAGATTTGCAAGATCTGCCCAGAGTCCAGCTGCTCATACTAG 4980  
Db 4921 AAGTCCCAATTAAGATTTGCAAGATCTGCCCAGAGTCCAGCTGCTCATACTAG 4980  
QY 4981 ATACTGCTGGGATTAATGCAATGGAAGAGTGTGAGCAGAGGAGCAGAAATTAAG 5040  
Db 4981 ATACTGCTGGGATTAATGCAATGGAAGAGTGTGAGCAGAGGAGCAGAAATTAAG 5040  
QY 5041 CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGTTGTGCTGAGCCAGCAAG 5100  
Db 5041 CTTCAACAGAAAGGCTCAACAAAGAAATGTCATGTTGTGCTGAGCCAGCAAG 5100  
QY 5101 AATTTATGCTGTGTACAAAGTTTGCAGAAACCAACATCACTTAATCTAATTA 5160  
Db 5101 AATTTATGCTGTGTGTACAAAGTTTGCAGAAACCAACATCACTTAATCTAATTA 5160  
QY 5161 CTGAAGAGACTACTCATGTTGTTATGAAGCAATGCTGAGTTGTGTGAAGGCAAC 5220  
Db 5161 CTGAAGAGACTACTCATGTTGTTATGAAGCAATGCTGAGTTGTGTGAAGGCAAC 5220  
QY 5221 TGAATATTTTCTAGCAATTTGCGGAGAGAAATGAGTATTTCTGAGTACCC 5280  
Db 5221 TGAATATTTTCTAGCAATTTGCGGAGAGAAATGAGTATTTCTGAGTACCC 5280  
QY 5281 AGTCTATTAAGAAAGAAATGCTGAATGAGCATGTTTGAAGTCAAGAGAGATGTTG 5340

Db	5281	AGTCTATTAAAGAAAGAAAATGCTGAATAGCATGATTTTGGAGTCAGAGAGAAATGTGG	5340
QY	5341	TCGAATGGAAGAAACCCACCAAGGTCCAAAGCGAGCAGAAGAATCCGAGACAGAAAGATCT	5400
Db	5341	TCGAATGGAAGAAACCCACCAAGGTCCAAAGCGAGCAGAAGAATCCGAGACAGAAAGATCT	5400
QY	5401	TCAGGGGGCTAGAAATTTGTTGCTATGAGGCCCTTACCAACATGCCACAGATCACTGG	5460
Db	5401	TCAGGGGGCTAGAAATTTGTTGCTATGAGGCCCTTACCAACATGCCACAGATCACTGG	5460
QY	5461	AATGGATGTCACAGCTGTGTGTCCTCTGTGTGTAAGAGACTTTCATTCATCCCTTG	5520
Db	5461	AATGGATGTCACAGCTGTGTGTCCTCTGTGTGTAAGAGACTTTCATTCATCCCTTG	5520
QY	5521	GCACAGGTGTCACCCCAATTGTGTTGTGACAGCCAGATGCCGTGCACAGAGACAATGGCT	5580
Db	5521	GCACAGGTGTCACCCCAATTGTGTTGTGACAGCCAGATGCCGTGCACAGAGACAATGGCT	5580
QY	5581	TCCATGCAATTTGGGCAATGTGTGAGGCACTGTGTGTGACCCCGAGAGTGGGTGTTGGACA	5640
Db	5581	TCCATGCAATTTGGGCAATGTGTGAGGCACTGTGTGTGACCCCGAGAGTGGGTGTTGGACA	5640
QY	5641	GTCGACACACTCCACATGCGCAGAGAGGTGACACACTCACTGATGCCCAAGATCCCCACA	5700
Db	5641	GTCGACACACTCCACATGCGCAGAGAGGTGACACACTCACTGATGCCCAAGATCCCCACA	5700
QY	5701	GCCCACTACTGA 5711	
Db	5701	GCCCACTACTGA 5711	

```

RESULT 8
US-09-911-904-127
Sequence 127, Application US/09911904
Publication No. US20030096234A1
GENERAL INFORMATION:
APPLICANT: Fair, Spencer B.
APPLICANT: Pickett, Gavin G.
APPLICANT: Neft, Robin Eileen
APPLICANT: Dunn, II, Robert Thomas
TITLE OF INVENTION: CAINE TOXICITY GENES
FILE REFERENCE: 400742000200
CURRENT APPLICATION NUMBER: US/09/911,904
CURRENT FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: US 60/220,057
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0.
SEQ ID NO 127
LENGTH: 499
TYPE: DNA
ORGANISM: Canis familiaris
US-09-911-904-127

```

Query Match	6.4%;	Score 364.6;	DB 9;	Length 499;
Best Local Similarity	-86.7%;	Pred. No. 1.9e-82;		
Matches 419; Conservative	0;	Mismatches 49;	Indels 15;	Gaps 1

QY	5229	TTTTAGGAATTCGGGGAGGAAATGGGTACTTATGACTTTCTTGGCTGACCCAGTCATT	5288
	1	TTTTTCTTGGGATTCGAGGAGGAAATGGGTACTTATGACTTTCTTGGCTGACCCAGTCATT <td>60</td>	60
QY	5289	AAAGAAAGAAAATATGCTGAATGAGCATGATTTTGTAAGTCAGAGAGATGTGGTCAAATGA	5344
	61	AAAGAAAGAAAAGATCTAGATGACATGATTTTGTAAGTCAGAGAGATGTGGTCAAATGA	120
QY	5349	AGAAACCAACCAAGTCCCAAGCGAGCAAGAGATTCACG-----GACAGA	5399
	121	AGAAATCACCAAGGCTCCGAGCGAGCAAGAAATCCGAGCAGAGATCCCAAGACAGA	180
QY	5394	AAGATCTTCAAGGGGGCTAAGAAATCTGTTCATATGGGCCCTTCACCAATATGCCACAAAT	5455
	181	AAGATCTTCAAGGGGGCTAAGAAATCTGTTCATATGGGCCCTTCACCAATATGCCACAAAT	240

OY	5454	CAACTGGAAATGAGTACACCTGTGTGTCCTCTGGTGGAAGAACCTTTCAATC	5513
Dd	241	CAATTAGAGTGGATGGTGCACCTCTGTGGGCTCTGTGGTGAAGAACCTTCTGTAATC	300
OY	5514	ACCCTTGGCACAGGTGTCCACCCAATTGTGTGTGACAGCCAGATGCCGTGACAGAGAC	5573
Dd	301	ACCCTCAGACAGGCGACCTCATCCAGTGTATCTCTGCAGCCGAGACGCTTGACAGAGAC	360
OY	5574	AATGGCTTCATGCATAATTGGSCAATGTGTGAGGACCTGTGGTAGCCCGAGAGTGGGTG	5633
Dd	361	AGTGGCTTCCATGTGGATTTGGSCAATGTGTGAGGACCTGTGGTGTACCCGAGAATGGGTA	420
OY	5634	TTGACAGATGTAGCACTCTACCAAGTGCAGAGAGCTGGAACACTTACCTGATACCCCAGATC	5693
Dd	421	CTGACAGATGTAGGCCCTCTACCAAGTGCAGAGAGCTGGAACACTTACCTGATACCCCAGATT	480
OY	5694	CCC 5696	
Dd	481	CCC 483	

RESULT 9  
US-09-864-761-4552  
Sequence 4552, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIORITY APPLICATION NUMBER: US 60/180,312  
PRIORITY FILING DATE: 2000-02-04  
PRIORITY APPLICATION NUMBER: US 60/207,456  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/632,366  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: GB 24263,6  
PRIORITY FILING DATE: 2000-10-04  
PRIORITY APPLICATION NUMBER: US 60/236,359  
PRIORITY FILING DATE: 2000-09-27  
PRIORITY APPLICATION NUMBER: PCT/US01/00666  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00667  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00664  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00669  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00665  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00668  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00663  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00662  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00661  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00670  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: US 60/234,667  
PRIORITY FILING DATE: 2000-09-21  
PRIORITY APPLICATION NUMBER: US 09/608,408  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 09/774,203  
PRIORITY FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117









; LENGTH: 121  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-818-875-658

Query Match 2.1%; Score 121; DB 9; Length 121;  
Best Local Similarity 100.0%; Pred. No. 8.6e-21;  
Matches 121: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 CTGGTAAAGTTCATTGGAAGAAAGAAATGATTATCTGCTCTCGGTTGAAGAG 150  
|  
Db 121 CTGGTAAAGTTCATTGGAAGAAAGAAATGATTATCTGCTCTCGGTTGAAGAG 62  
  
QY 151 TACAAATGTCATTATGCTATGACAAAATCTTAGAGTGTCCATCTGCTGAGTTGA 210  
|  
Db 61 TACAAATGTCATTATGCTATGACAAAATCTTAGAGTGTCCATCTGCTGAGTTGA 2  
  
QY 211 T 211  
|  
Db 1 T 1

## RESULT 15

US-09-818-875-661  
; Sequence 661, Application US/09818875  
; Publication No. US20030051270A1  
; GENERAL INFORMATION:  
; APPLICANT: Kmiec, Eric B.  
; APPLICANT: Gamper, Howard B.  
; APPLICANT: Rice, Michael C.  
; TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
; FILE REFERENCE: Napro-4  
; CURRENT APPLICATION NUMBER: US/09/818,875  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: US 60/192,176  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: US 60/192,179  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: US 60/208,538  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: US 60/244,989  
; PRIOR FILING DATE: 2000-10-30  
; NUMBER OF SEQ ID NOS: 4385  
; SOFTWARE: Friedman macro Napro4  
; SEQ ID NO 661  
; LENGTH: 121  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-818-875-661

Query Match 2.1%; Score 121; DB 9; Length 121;  
Best Local Similarity 100.0%; Pred. No. 8.6e-21;  
Matches 121: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 ATGATTTATCTGCTCTTGGCTTGAGAGTACAAAATGCTTAATGCTATGCAGAAA 179  
|  
Db 1 ATGATTTATCTGCTCTTGGCTTGAGAGTACAAAATGCTTAATGCTATGCAGAAA 60  
  
QY 180 ATCTTAGAGTGTCCATCTGCTGAGATTGATCAAGAACTGCTCCACAAAGTGTGAC 239  
|  
Db 61 ATCTTAGAGTGTCCATCTGCTGAGATTGATCAAGAACTGCTCCACAAAGTGTGAC 120  
  
QY 240 C 240  
|  
Db 121 C 121

Search completed: June 27, 2003, 05:24:49  
Job time : 763.849 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 12:33:08 ; Search time 1078.23 Seconds  
(without alignments)  
11673.202 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_120\_5708

Perfect score: 5589  
Sequence: 1 ATGCAATTATCTGCTCTTCG.....AGATCCCGACAGCAGCACTAC 5589

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

```
1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5587.4	100.0	5711	19	AAV46449
2	5587.4	100.0	5711	21	AAC60795
3	5587.4	100.0	5711	17	AAT17493
4	5587.4	100.0	5711	17	AAT17495
5	5587	100.0	5711	19	AAV46463
6	5587	100.0	5711	19	AAV46464
7	5587	100.0	5711	19	AAV46458
8	5587	100.0	5711	19	AAV46459
9	5587	100.0	5711	19	AAV46460

10	5587	100.0	5711	19	AAV46461	Human BRCA1 omi2 p
11	5587	100.0	5711	19	AAV46462	Human BRCA1 omi2 p
12	5585.8	99.9	5712	18	AAT84840	Human breast and o
13	5585.8	99.9	5712	21	AAZ87995	BRCA1 gene sequenc
14	5585.8	99.9	5914	17	AAT32601	BRCA1, breast and
15	5585.8	99.9	5914	17	AAT17438	BRCA1 coding seque
16	5585.8	99.9	5914	17	AAT18310	BRCA1, breast and
17	5585.8	99.9	14533	23	AAZ5883	DNA encoding novel
18	5584.2	99.9	5711	17	AAT42024	BRCA1 allele #8403
19	5584.2	99.9	5711	17	AAT42030	BRCA1 allele #7408
20	5584.2	99.9	5711	18	AAT70067	Mutant BRCA1 allel
21	5584.2	99.9	5711	18	AAT70073	Mutant BRCA1 allel
22	5584.2	99.9	5711	19	AAV60569	Tumorigenic BRCA1
23	5584.2	99.9	5711	19	AAV60569	Tumorigenic BRCA1
24	5584.2	99.9	5914	17	AAT17445	Mutated BRCA1 codi
25	5584.2	99.9	5914	17	AAT17446	Mutated BRCA1 codi
26	5584.2	99.9	5914	17	AAT17449	Mutated BRCA1 codi
27	5584.2	99.9	5914	17	AAT17453	Mutated BRCA1 codi
28	5584.2	99.9	5914	17	AAT17457	Mutated BRCA1 codi
29	5584.2	99.9	5914	17	AAT17462	Mutated BRCA1 codi
30	5584.2	99.9	5914	17	AAT17466	Mutated BRCA1 codi
31	5584.2	99.9	5914	17	AAT17468	Mutated BRCA1 codi
32	5584.2	99.9	5914	17	AAT17471	Mutated BRCA1 codi
33	5584.2	99.9	5914	17	AAT17476	Mutated BRCA1 codi
34	5584.2	99.9	5914	17	AAT17478	Mutated BRCA1 codi
35	5584.2	99.9	5914	17	AAT17479	Mutated BRCA1 codi
36	5584.2	99.9	5914	17	AAT17481	Mutated BRCA1 codi
37	5584.2	99.9	5914	17	AAT17483	Mutated BRCA1 codi
38	5584.2	99.9	5914	17	AAT17484	Mutated BRCA1 codi
39	5584.2	99.9	5914	17	AAT17486	Mutated BRCA1 codi
40	5584.2	99.9	5914	17	AAT17489	Mutated BRCA1 codi
41	5584.2	99.9	5914	17	AAT17490	Mutated BRCA1 codi
42	5584.2	99.9	5914	17	AAT17491	Mutated BRCA1 codi
43	5584.2	99.9	5914	17	AAT17492	Mutated BRCA1 codi
44	5584.2	99.9	5914	17	AAT17494	Mutated BRCA1 codi
45	5584.2	99.9	5914	17	AAT17496	Mutated BRCA1 codi

#### ALIGNMENTS

RESULT 1	AAV46449	standard; cDNA; 5711 BP.
ID	AAV46449	
AC	AAV46449	
XX	18-NOV-1998	(first entry)
XX	Human BRCA1 omi2 cDNA.	
DE	Human BRCA1 omi2 cDNA.	
XX	BRCA1; omi2; human; breast and ovarian cancer predisposing gene;	
KW	polymorphism; susceptibility; anti-oncogene; tumour suppressor;	
KW	chromosome 17q; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	120..5711
FT		/*tag= a
FT		/product= "BRCA1 omi2 protein"
XX		
PN	US5750400-A.	
XX		
PD	12-MAY-1998.	
XX		
PF	12-FEB-1997;	97US-0798691.
XX		
PR	12-FEB-1996;	96US-0598591.
XX	12-FEB-1997;	97US-0798691.
XX		
PA	(ONCO-) ONCORMED INC.	
XX		

PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Schelter DB, Zeng B;  
XX  
XX MPI: 1998-296774/26.  
DR P-PSDB: AAM76099.  
PT BRCA1 om1 gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer  
XX  
PS Claim 2d; Column 41-46; 54pp; English.  
XX  
XX This sequence encodes the human BRCA1 (breast and ovarian cancer  
CC predisposing gene) om2 gene. This sequence and polymorphic variations of  
CC this sequence are useful for the identification of an individual who may  
CC or may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers,  
CC especially breast and ovarian cancer. It is found at human chromosome 17q  
CC which is known to be linked to cancer susceptibility, especially breast  
CC cancer. Cells containing a mutation in this gene lose the wild-type  
CC function of BRCA1 and are more susceptible to cancers.  
XX  
XX

Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1383 T; 0 other;

Query Match 100.0%; Score 5587.4; DB 19; Length 5711;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGATTTATGCTCTCTGCGGTGAGAGTACAAATGTCTAATGCTATGAGAAA 60  
DB 120 ATGATTTATGCTCTCTGCGGTGAGAGTACAAATGTCTAATGCTATGAGAAA 179  
OY 61 ATCTAGAGTGTCCATCTGCTGAGTGTATCAAGAACCTGTCCCAAAAGTGTGAC 120  
DB 180 ATCTAGAGTGTCCATCTGCTGAGTGTATCAAGAACCTGTCCCAAAAGTGTGAC 239  
OY 121 CACATATTTTGGCAATTTTGCATGCTGAAACCTTCTACACAGAGAGAGGCTTCACAG 180  
DB 240 CACATATTTTGGCAATTTTGCATGCTGAAACCTTCTACACAGAGAGAGGCTTCACAG 299  
OY 181 TGTCTTTATGTAAGATATATATACCAAGAGGCTTACAGAAAGTACAGATTTAGT 240  
DB 300 TGTCTTTATGTAAGATATATATACCAAGAGGCTTACAGAAAGTACAGATTTAGT 359  
OY 241 CAACCTGTGAGAGCTATTTGAAATCATTTGCTTTACAGTGTACACAGGTTGGAG 300  
DB 360 CAACCTGTGAGAGCTATTTGAAATCATTTGCTTTACAGTGTACACAGGTTGGAG 419  
OY 301 TATGCAACAGCTATATTTTGCACAAAAAGAAATACCTCTCTGACATCTAAAGAT 360  
DB 420 TATGCAACAGCTATATTTTGCACAAAAAGAAATACCTCTCTGACATCTAAAGAT 479  
OY 361 GAGATTTTATCATCCAAAGTATGGGCTTACAGAAACCGTCCAAAAGACTTCTACAGAT 420  
DB 480 GAGATTTTATCATCCAAAGTATGGGCTTACAGAAACCGTCCAAAAGACTTCTACAGAT 539  
OY 421 GAACCCGAAATCTCTCTCTGAGAGAAACAGCTCAGTGTCCAACTCTTAACCTTGA 480  
DB 540 GAACCCGAAATCTCTCTCTGAGAGAAACAGCTCAGTGTCCAACTCTTAACCTTGA 599  
OY 481 ACTGTGAGAACTCTGAGAGCAAGAGCGGATACAACTCAAAAGAGCTGTCTACATT 540  
DB 600 ACTGTGAGAACTCTGAGAGCAAGAGCGGATACAACTCAAAAGAGCTGTCTACATT 659  
OY 541 GAATGGGATCTGATTTCTTCTGAAGATACCGTTAATAGGCAACTTATGCACTGGGA 600  
DB 660 GAATGGGATCTGATTTCTTCTGAAGATACCGTTAATAGGCAACTTATGCACTGGGA 719  
OY 601 GATCAAGATTTGTTCAAAATCACCCTCAAGAGACGAGGATGAATCATGTTGATTT 660  
DB 720 GATCAAGATTTGTTCAAAATCACCCTCAAGAGACGAGGATGAATCATGTTGATTT 779

OY 661 GCAAAAAGGCTCTTGTGAATTTTCTGAGAGGATGTAAACAATACTGAACATCATCA 720  
DB 780 GCAAAAAGGCTCTTGTGAATTTTCTGAGAGGATGTAAACAATACTGAACATCATCA 839  
OY 721 CCCAGTAATATGATTTTGAACACCACTGAGAGGCTGAGAGGCTATCCAGAAAAG 780  
DB 840 CCCAGTAATATGATTTTGAACACCACTGAGAGGCTGAGAGGCTATCCAGAAAAG 899  
OY 781 TATCAGGATGTTCTTTTCAAACTGTGAGAGCCATGTGGCAAAATACTATGCC 840  
DB 900 TATCAGGATGTTCTTTTCAAACTGTGAGAGCCATGTGGCAAAATACTATGCC 959  
OY 841 AGCTATTTACAGATGAGAGACGAGCTTATATCTACTAATAAGAGATGATGTAGAA 900  
DB 960 AGCTATTTACAGATGAGAGACGAGCTTATATCTACTAATAAGAGATGATGTAGAA 1019  
OY 901 AAGCTGAATTTCTGTATATAAAGCAAAACGCTGCTTACAGAGAGGCCAATATACGA 960  
DB 1020 AAGCTGAATTTCTGTATATAAAGCAAAACGCTGCTTACAGAGAGGCCAATATACGA 1079  
OY 961 TGGGCTGGAAGTAAAGAAACATGTATGTATAGGCGGACTCCAGCAGAAAAAGTA 1020  
DB 1080 TGGGCTGGAAGTAAAGAAACATGTATGTATAGGCGGACTCCAGCAGAAAAAGTA 1139  
OY 1021 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGGAATATACAGAACTGCCATGC 1080  
DB 1140 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGGAATATACAGAACTGCCATGC 1139  
OY 1081 TCAGAGAACTCTAGAGATCTAGAGATGTCTCTTGTGATATACCTAATATAGCAGCTTCAG 1140  
DB 1200 TCAGAGAACTCTAGAGATCTAGAGATGTCTCTTGTGATATACCTAATATAGCAGCTTCAG 1259  
OY 1141 AAGTAAATAGAGGTTTCCAGAGTGTGAGAGTGTAGGTTGATGAGCTCATCATGAT 1200  
DB 1260 AAGTAAATAGAGGTTTCCAGAGTGTGAGAGTGTAGGTTGATGAGCTCATCATGAT 1319  
OY 1201 GGGGAGTCTGAATCAAAATCCCAAGTAGCTGATGATTTGAGAGCTTCTAATATAGGTAGAT 1260  
DB 1320 GGGGAGTCTGAATCAAAATCCCAAGTAGCTGATGATTTGAGAGCTTCTAATATAGGTAGAT 1379  
OY 1261 GAATATTTCTGCTTCTTACAGAGAAATAGACTTACTGGCCAGTGTATCTCATGAGGCTTTA 1320  
DB 1380 GAATATTTCTGCTTCTTACAGAGAAATAGACTTACTGGCCAGTGTATCTCATGAGGCTTTA 1439  
OY 1321 ATATGTAAAGTGAAGAGTGTCTACTCCAAATAGTAGAGATTAATGAAGACAAAATA 1380  
DB 1440 ATATGTAAAGTGAAGAGTGTCTACTCCAAATAGTAGAGATTAATGAAGACAAAATA 1499  
OY 1381 TTTGGAAACCTATCGAAGAGGCAAGCCCTCCCAACTTAAGCATGTAACTGAATAAT 1440  
DB 1500 TTTGGAAACCTATCGAAGAGGCAAGCCCTCCCAACTTAAGCATGTAACTGAATAAT 1559  
OY 1441 CTATATTTAGAGACTTTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCAAAAT 1500  
DB 1560 CTATATTTAGAGACTTTTGTACTGAGCCACAGATTAATACAGAGGCTCCCTCAAAAT 1619  
OY 1501 AAATTAACCTTAAGAGAGACCTACATCAGGCTTCACTCCGAGGATTTTATCAAGAAA 1560  
DB 1620 AAATTAACCTTAAGAGAGACCTACATCAGGCTTCACTCCGAGGATTTTATCAAGAAA 1679  
OY 1561 GCAGATTTTGGAGTTCAAAAAGACTCTGAAATGATTAATACAGAGAACTTAACCAACGAG 1620  
DB 1680 GCAGATTTTGGAGTTCAAAAAGACTCTGAAATGATTAATACAGAGAACTTAACCAACGAG 1739  
OY 1621 CAGATTTGCTCAAGTATGATATTTACTAATAGTGTCTATGAGATTAACCAAGGATGAT 1680  
DB 1740 CAGATTTGCTCAAGTATGATATTTACTAATAGTGTCTATGAGATTAACCAAGGATGAT 1799  
OY 1681 TCTATTTACAGATGAGAAAAATCTTAACCAATAGATACAGAAAAAGATTCGCTTTC 1740  
DB 1800 TCTATTTACAGATGAGAAAAATCTTAACCAATAGATACAGAAAAAGATTCGCTTTC 1859

1741 AAAACGAAAGCTGAACCTATTAAGCAGAGCTATTAAGCAATATGGAACCTGAACTTAATTAATATC 1800  
1860 AAAAGAAAGCTGAACCTATTAAGCAGAGCTATTAAGCAATATGGAACCTGAACTTAATTAATATC 1919  
1801 CACAAATTCAAAAGCAGCTTAAAAAGATAGGCTGAGAGAGAGCTCTTCTACAGGCAATAT 1860  
1920 CACAAATTCAAAAGCAGCTTAAAAAGATAGGCTGAGAGAGAGCTCTTCTACAGGCAATAT 1979  
1861 CATGGCTTGAAGCTAGTACGATAGAAATCTAAGCCCACTAATTTGTTCTGTAATTTGCA 1920  
1980 CATGGCTTGAAGCTAGTACGATAGAAATCTAAGCCCACTAATTTGTTCTGTAATTTGCA 2039  
1921 ATTGATAGTGTCTAGCAGTGAAGAGATTAAGAAAAAGTACAACTAATCCCAATCCAGTC 1980  
2040 ATTGATAGTGTCTAGCAGTGAAGAGATTAAGAAAAAGTACAACTAATCCCAATCCAGTC 2099  
1981 AGGCACAGCAGAAAGCTTACAACTCATGTAAGGTAAGAACTGCAACTGGAGCCAGAAAG 2040  
2100 AGGCACAGCAGAAAGCTTACAACTCATGTAAGGTAAGAACTGCAACTGGAGCCAGAAAG 2159  
2041 AGTAACAGCCAAATGAGACAGACAAATGAAGACATGAGTATCTTCCAGAGCTG 2100  
2160 AGTAACAGCCAAATGAGACAGACAAATGAAGACATGAGTATCTTCCAGAGCTG 2219  
2101 AAGTTAACAATGACACCTGTTCTTTTACTAAGTGTTCAAATACAGTAACTTAAGAA 2160  
2220 AAGTTAACAATGACACCTGTTCTTTTACTAAGTGTTCAAATACAGTAACTTAAGAA 2279  
2161 TTTGTCAATCTAGCTTCCAG 2220  
2280 TTTGTCAATCTAGCTTCCAG 2339  
2221 TCTAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGAGAGAGAGAGAGAG 2280  
2340 TCTAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGAGAGAGAGAGAGAG 2399  
2281 GAAGATCTGTAGAGAGAGAGAGATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2340  
2400 GAAGATCTGTAGAGAGAGAGATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2459  
2341 GAAGATCTGTAGAGAGAGAGATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2400  
2460 GAAGATCTGTAGAGAGAGAGATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2519  
2401 TGTGTAGTCACTGTGACAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
2520 TGTGTAGTCACTGTGACAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2579  
2461 GATTAATAGAAATGACACAGAGAGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGT 2520  
2580 GATTAATAGAAATGACACAGAGAGCTTTAAGTATCCATTGGGACATGAAGTTAACCAAGT 2639  
2521 CGGGAAGACAGATGAAATGGAAGAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2580  
2640 CGGGAAGACAGATGAAATGGAAGAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2699  
2581 TTTCAAGGTTTCAAGAGAGAGAGATTTGCTGTTTCAATTCAGAGAGAGAGAGAGAGAG 2640  
2700 TTTCAAGGTTTCAAGAGAGAGAGATTTGCTGTTTCAATTCAGAGAGAGAGAGAGAGAG 2759  
2641 GAATGTGCAACATTTCTGCCCCAGCTGCGGCTTTAAGAGAGAGAGAGAGAGAGAGAGAG 2700  
2760 GAATGTGCAACATTTCTGCCCCAGCTGCGGCTTTAAGAGAGAGAGAGAGAGAGAGAGAG 2819  
2701 TTTGATGATGACAAAG 2760  
2820 TTTGATGATGACAAAG 2879  
2761 CAGACAGTAAATATCACTGACAGGCTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2820  
2880 CAGACAGTAAATATCACTGACAGGCTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2939  
2821 AATGCCAAATGTATGATCAAGAGAGAGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2880

2940 AATGCCAAATGTATGATCAAGAGAGAGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2999  
2881 AATGCCAAATGTATGATCAAGAGAGAGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2940  
3000 AATGCCAAATGTATGATCAAGAGAGAGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3059  
2941 CCACACTTTTCCATCAAGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3000  
3060 CCACACTTTTCCATCAAGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3119  
3001 GAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060  
3120 GAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3179  
3061 AGTACAGTGAAGCAATTTAGCCCTAATTAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 3120  
3180 AGTACAGTGAAGCAATTTAGCCCTAATTAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 3239  
3121 TCAGCAATATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3180  
3240 TCAGCAATATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3299  
3181 ATAGGTTCCAGTATGAG 3240  
3300 ATAGGTTCCAGTATGAG 3359  
3241 AATGCTATGCTTGAATTTAGGAGGTTTTCAGACCTGAGGCTATTAACAAAGTCTTCTGCA 3300  
3360 AATGCTATGCTTGAATTTAGGAGGTTTTCAGACCTGAGGCTATTAACAAAGTCTTCTGCA 3419  
3301 AGTAATTTGATGAGTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3360  
3420 AGTAATTTGATGAGTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3479  
3361 AATGAGATTTCTCTCATATCTGATTTAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3420  
3480 AATGAGATTTCTCTCATATCTGATTTAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3539  
3421 CATGATCTCAGGTTTGTCTGAGACACCTGATGAGGCTGTTAATGATGATGATGATGATG 3480  
3540 CATGATCTCAGGTTTGTCTGAGACACCTGATGAGGCTGTTAATGATGATGATGATGATG 3599  
3481 GAAGATCTAGTATTTCTGCTGAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3540  
3600 GAAGATCTAGTATTTCTGCTGAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3659  
3541 GTCCAGAAAGAGAGAGCTTGAAGAGAGTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3600  
3660 GTCCAGAAAGAGAGAGCTTGAAGAGAGTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3719  
3601 GGTTCAGGAAAGAGAGAGCTTGAAGAGAGTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3660  
3720 GGTTCAGGAAAGAGAGAGCTTGAAGAGAGTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3779  
3661 GAAGAGCTTCCCTGCTTCCAACTGTTATTTGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3720  
3780 GAAGAGCTTCCCTGCTTCCAACTGTTATTTGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3839  
3721 TCTACTAGGCAATGACAGGCTTGTACAGAGTGTCTGTAGAGACAGAGAGAGAGAGAGAG 3780  
3840 TCTACTAGGCAATGACAGGCTTGTACAGAGTGTCTGTAGAGACAGAGAGAGAGAGAGAG 3899  
3781 TTTATCTTAAGAAATGCTTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3840  
3900 TTTATCTTAAGAAATGCTTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3959  
3841 CAGGAACATCACCCTTATGAG 3900  
3960 CAGGAACATCACCCTTATGAG 4019  
3901 AGTGAATTTGAAGAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3960

```

Db 4020 AGTGAATTGGAGACTTGAAGTGAACAAACACCCAGATCCCTTTTGATTGGTTCT 4079
OY 3961 TCCAAACAATGAGCATCATCTGAAGCCAGGAGTTGCTGATGATGACAAAGAAATTG 4020
Db 4080 TCCAAACAATGAGCATCATCTGAAGCCAGGAGTTGCTGATGATGACAAAGAAATTG 4139
OY 4021 GTTTCAGATGATGAGAAAGAGAGAGGCGCTTGGAGAAATATCAAGAAGACAAAGC 4080
Db 4140 GTTTCAGATGATGAGAAAGAGAGAGGCGCTTGGAGAAATATCAAGAAGACAAAGC 4199
OY 4081 ATGATTAACCTTAGTGAAGAGCATCTGGGTGAGAGTGAACCAACCCGCTCTGAA 4140
Db 4200 ATGATTAACCTTAGTGAAGAGCATCTGGGTGAGAGTGAACCAACCCGCTCTGAA 4259
OY 4141 GACTGCTCAGGCTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATG 4200
Db 4260 GACTGCTCAGGCTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATG 4319
OY 4201 CAACATTAACCTTAAGTCTCAGAGAGGAAATGGCTGAACCTAGAGCTGTGTAGAACAG 4260
Db 4320 CAACATTAACCTTAAGTCTCAGAGAGGAAATGGCTGAACCTAGAGCTGTGTAGAACAG 4379
OY 4261 CATGGAGGCGAGCTTCTTAACAGCTACCTTCATCATTAAGTACTCTTGCCCTTGAG 4320
Db 4380 CATGGAGGCGAGCTTCTTAACAGCTACCTTCATCATTAAGTACTCTTGCCCTTGAG 4439
OY 4321 GACCTGCGAAATCCAGAACAAAGCACATCAGAAAAGAGATTTAACTTACAGAAAGAT 4380
Db 4440 GACCTGCGAAATCCAGAACAAAGCACATCAGAAAAGAGATTTAACTTACAGAAAGAT 4499
OY 4381 AGTGAATACCTTAAGTGAAGCAATCCAGAAAGCCTTCTGCTGACAAAGTTGAGGTCT 4440
Db 4500 AGTGAATACCTTAAGTGAAGCAATCCAGAAAGCCTTCTGCTGACAAAGTTGAGGTCT 4559
OY 4441 GCAGATAGTTTACCAAGTAAATAAGAAACAGAGAGTGAAGAGTCAATCCCTTTTAA 4500
Db 4560 GCAGATAGTTTACCAAGTAAATAAGAAACAGAGAGTGAAGAGTCAATCCCTTTTAA 4619
OY 4501 TGCCCATCATTAAGATAGTGTGATGACAGTGTCTGAGAGTCTTGAGAAATAG 4560
Db 4620 TGCCCATCATTAAGATAGTGTGATGACAGTGTCTGAGAGTCTTGAGAAATAG 4679
OY 4561 AACTACCATCTCAGAGAGAGTCAATTAAGTGTGATGATGAGAGAGCAACAGCTGAA 4620
Db 4680 AACTACCATCTCAGAGAGAGTCAATTAAGTGTGATGATGAGAGAGCAACAGCTGAA 4739
OY 4621 GAGTCTGGGCGACAGATTTGAGGAAACATCTTACTTCCAGAGCAAGATCTAGAGGA 4680
Db 4740 GAGTCTGGGCGACAGATTTGAGGAAACATCTTACTTCCAGAGCAAGATCTAGAGGA 4799
OY 4681 ACCCTTAACCTGGAATCTGGAATCAGCCTCTCTGATGACCTGGAATCTGATCCTCT 4740
Db 4800 ACCCTTAACCTGGAATCTGGAATCAGCCTCTCTGATGACCTGGAATCTGATCCTCT 4859
OY 4741 GAAAGCAGAGGCGGAGAGTGTGTGCAACATCCATTTCAACCTCTGATG 4800
Db 4860 GAAAGCAGAGGCGGAGAGTGTGTGCAACATCCATTTCAACCTCTGATG 4919
OY 4801 AAAGTTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCTATACTACT 4860
Db 4920 AAAGTTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCTATACTACT 4979
OY 4861 GATACCTGCTGGGTATATGCAATGGAAGAAAGTGTGAGCAGAGGAGAGCCAGAAATTGACA 4920
Db 4980 GATACCTGCTGGGTATATGCAATGGAAGAAAGTGTGAGCAGAGGAGAGCCAGAAATTGACA 5039
OY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGAGTGTGTGAGGCTGACCCAGAA 4980
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGAGTGTGTGAGGCTGACCCAGAA 5099
OY 4981 GAATTTATGCTGCTGATACAAATTTGCGAGAAACCCAGATCTTAACTTAATCAAT 5040
Db 5100 GAATTTATGCTGCTGATACAAATTTGCGAGAAACCCAGATCTTAACTTAATCAAT 5159

```

```

OY 5041 ACTGAGAGACTACATGTTTATGAAACAGATGCTGAGTTTGTGTGAGAGACA 5100
Db 5160 ACTGAGAGACTACATGTTTATGAAACAGATGCTGAGTTTGTGTGAGAGACA 5219
OY 5101 CTGAATATTTTCTAGGAATTTGGGAGGAGAAATGGGTAGTATTTCTGGGTGAC 5160
Db 5220 CTGAATATTTTCTAGGAATTTGGGAGGAGAAATGGGTAGTATTTCTGGGTGAC 5279
OY 5161 CACTCTATTAAAGAAAGAAATGCTGATGACATGATTTTGAAGTCAAGAGAGATGTG 5220
Db 5280 CACTCTATTAAAGAAAGAAATGCTGATGACATGATTTTGAAGTCAAGAGAGATGTG 5339
OY 5221 GTCAATGGAAGAAACCAACAGGTCGAAAGCAGAGAGAAATCCAGAGAGAAAGATC 5280
Db 5340 GTCAATGGAAGAAACCAACAGGTCGAAAGCAGAGAGAAATCCAGAGAGAAAGATC 5399
OY 5281 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCGCTTACCAACATGCCAGATCACTG 5340
Db 5400 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCGCTTACCAACATGCCAGATCACTG 5459
OY 5341 GAATGATGATACAGCTGTGTGCTCTCTGATGAGAGAGAGTTCATTCATCACTT 5400
Db 5460 GAATGATGATACAGCTGTGTGCTCTCTGATGAGAGAGAGTTCATTCATCACTT 5519
OY 5401 GGCACAGTGTCCACCAATTTGTTGTGACAGCAGATGCTGAGACAGAGACAAATGAC 5460
Db 5520 GGCACAGTGTCCACCAATTTGTTGTGACAGCAGATGCTGAGACAGAGACAAATGAC 5579
OY 5461 TTCATGCAATTTGGGAGATGTGTAGAGCACCCTGTGTGACCGAGATGGGTGTGAC 5520
Db 5580 TTCATGCAATTTGGGAGATGTGTAGAGCACCCTGTGTGACCGAGATGGGTGTGAC 5639
OY 5521 AGTGTAGCACTTACAGAGCCAGAGGTCGACACTACCTGATATCCCAATCCCAAT 5580
Db 5640 AGTGTAGCACTTACAGAGCCAGAGGTCGACACTACCTGATATCCCAATCCCAAT 5699
OY 5581 AGCCACTAC 5589
Db 5700 AGCCACTAC 5708

RESULT 2
AAC60795
AAC60795 standard; cDNA: 5711 BP.
AC
AAC60795;
XX
DT 07-FEB-2001 (first entry)
XX
DE Human BRCA1 (om13) nucleotide sequence SEQ ID NO:5.
XX
KW Human: BRCA1: chromosome 17: 17q21; breast cancer; ovarian cancer;
KW gene therapy; diagnosis; cytostatic; genetic susceptibility; mutation;
KW polymorphism; identification; ss.
XX
OS Homo sapiens.
XX
PN US6130322-A.
PD 10-OCT-2000.
XX
PF 06-MAY-1998; 98US-0074476.
XX
PR 12-FEB-1996; 96US-0598591.
PR 12-DEC-1997; 97US-0798691.
XX
PA (GENE-) GENE LOGIC INC.
PI Zeng B, Thurber D, Olson SJ, Alvares CP, Allen ACP, Murphy PD;
PI Critz BS;
XX
DR WPI: 2000-646756/62.

```

DR P-PSDB; AAB24219.

XX New coding sequence of the human BRCA1 gene, i.e. BRCA1 (om12), useful  
PT in gene therapy, especially for preventing or treating breast or  
PT ovarian cancer, as well as for diagnosing or monitoring breast or  
PT ovarian cancer

PS Example 4; Column 59-66; 56pp; English.

XX AAC60793 to AAC60795 encode the human BRCA1 (om11-3) proteins given in  
CC AAB24217 to AAB24219, respectively. BRCA1 is found on chromosome 17  
CC mapping to position 17q21. The BRCA1 (om12) coding sequence is  
CC specifically claimed in the present invention. The BRCA1 (om12) coding  
CC sequence is useful in gene therapy, especially for preventing or treating  
CC breast or ovarian cancer. It is also useful for diagnosing or monitoring  
CC breast or ovarian cancer. Furthermore, the BRCA1 (om12) coding sequence  
CC is useful for: (a) identifying individuals having BRCA1 gene mutations  
CC and having an increased genetic susceptibility to breast or ovarian  
CC cancer, or identifying a mutation that increases the genetic  
CC susceptibility to breast or ovarian cancer; (b) avoiding  
CC misinterpretation of polymorphisms found in the BRCA1 gene; (c)  
CC determining the presence of a previously unknown mutation in the BRCA1  
CC gene; (d) probing a human sample of the BRCA1 gene by allele to determine  
CC the presence of either polymorphic alleles or mutations; and (e)  
CC performing diagnosis with a reagent derived from the BRCA1 (om1) cDNA  
CC sequence. AAC60796 to AAC60861 represent PCR primers for the BRCA1 gene,  
CC which are used in an example from the present invention.

SO Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1383 T; 0 other:

Query Match 100.0%; Score 5587.4; DB 21; Length 5711;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGATTATCTGCTCTTGGCGTTGAAGAATGCAAAATGCTATATGCTATGCGAAA 60  
DB 120 ATGGGATTATCTGCTCTTGGCGTTGAAGAATGCAAAATGCTATATGCTATGCGAAA 179  
QY 61 ATCTTGAAGTCCCTGCTGTGTGAGTTCAGTCAAGAACCTGCTCCCAAAAGTGTGAC 120  
DB 180 ATCTTGAAGTCCCTGCTGTGTGAGTTCAGTCAAGAACCTGCTCCCAAAAGTGTGAC 239  
QY 121 CACATATTTTGCATTTTGCATGCTGAATCTTCAACAGAAAGAGGCTTTCACAG 180  
DB 240 CACATATTTTGCATTTTGCATGCTGAATCTTCAACAGAAAGAGGCTTTCACAG 299  
QY 181 TGTCTTATTTGTAAGATATATACCAAAAGAGCTTACAGAAAGTGTGAC 240  
DB 300 TGTCTTATTTGTAAGATATATACCAAAAGAGCTTACAGAAAGTGTGAC 359  
QY 241 CAACTGTGAGAGCTATTTGAAATCATTTGCTTTACAGCTTACACAGGTTGGAG 300  
DB 360 CAACTGTGAGAGCTATTTGAAATCATTTGCTTTACAGCTTACACAGGTTGGAG 419  
QY 301 TATGCAACAGCTATATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCAT 360  
DB 420 TATGCAACAGCTATATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCAT 479  
QY 361 GAAGTTTCTATCATCAAAAGTATGAGCTACAGAAACCGTCCAAAAGACTTTACAGAGT 420  
DB 480 GAAGTTTCTATCATCAAAAGTATGAGCTACAGAAACCGTCCAAAAGACTTTACAGAGT 539  
QY 421 GAACCCGAAATCTCTTCTGAGAAACAGCTCAGTGTCCAACTCTCTAACTTTGGA 480  
DB 540 GAACCCGAAATCTCTTCTGAGAAACAGCTCAGTGTCCAACTCTCTAACTTTGGA 599  
QY 481 ACTGTGAGAACTCTGAGCAAAAGCGGAGATCAACCTCAAAAAGAGCTCTGTCTACATT 540  
DB 600 ACTGTGAGAACTCTGAGCAAAAGCGGAGATCAACCTCAAAAAGAGCTCTGTCTACATT 659  
QY 541 GAATTTGGATCTGATCTCTGAAAGTACCGTTAATTAAGCAACTATTGCGAGTGGGA 600  
DB 660 GAATTTGGATCTGATCTCTGAAAGTACCGTTAATTAAGCAACTATTGCGAGTGGGA 719

QY 601 GATCAGAAATTTTACAAATCACCCCTCAAGAGAGGATGAATACATTGGATTCT 660  
DB 720 GATCAGAAATTTTACAAATCACCCCTCAAGAGAGGATGAATACATTGGATTCT 779  
QY 661 GCAAAAAAGGCTCTTGTGAATTTTCTGAGACGATGTAACAAAATCTGAACATCAA 720  
DB 780 GCAAAAAAGGCTCTTGTGAATTTTCTGAGACGATGTAACAAAATCTGAACATCAA 839  
QY 721 CCCAGTAATATATTTTGAACACCACTGAGAACCGTGCACCTAGAGGCTCCAGAAAAG 780  
DB 840 CCCAGTAATATATTTTGAACACCACTGAGAACCGTGCACCTAGAGGCTCCAGAAAAG 899  
QY 781 TATCAGGATGATTTCTGTTTCAAACTTGCAATGTGAGGACATGTGGCAAAATCTCATGCC 840  
DB 900 TATCAGGATGATTTCTGTTTCAAACTTGCAATGTGAGGACATGTGGCAAAATCTCATGCC 959  
QY 841 AGCTCATTTACAGCATGAGAACACAGCTTTATTTACTCATAAAGACAGAAATGAATAGAA 900  
DB 960 AGCTCATTTACAGCATGAGAACACAGCTTTATTTACTCATAAAGACAGAAATGAATAGAA 1019  
QY 901 AAGGCTGAATTTCTGTATTAATAAGCAACAGCTGCTTACCAAGAGCCACATTAACAGA 960  
DB 1020 AAGGCTGAATTTCTGTATTAATAAAGCAACAGCTGCTTACCAAGAGCCACATTAACAGA 1079  
QY 961 TGGGCTGGAAGTAAAGAAACATGTATGATGAGGCGAGCTCCAGCAGAAAAAAGGTA 1020  
DB 1080 TGGGCTGGAAGTAAAGAAACATGTATGATGAGGCGAGCTCCAGCAGAAAAAAGGTA 1139  
QY 1021 GATCTGAATCTGATCCCTGCTGTGTGAGAGAAAAAGAAATGAATTAAGCAGAACTGCCATGC 1080  
DB 1140 GATCTGAATCTGATCCCTGCTGTGTGAGAGAAAAAGAAATGAATTAAGCAGAACTGCCATGC 1199  
QY 1081 TCAGAGAAATCTTGAATATCTGAAGATGTTCTTGTGATTAACACTAATAGACATTTGAC 1140  
DB 1200 TCAGAGAAATCTTGAATATCTGAAGATGTTCTTGTGATTAACACTAATAGACATTTGAC 1259  
QY 1141 AAGGTAATGAGAGGTTTTCACAGAGTGAATCTTGAAGTCTGATGCTACATGAT 1200  
DB 1260 AAGGTAATGAGAGGTTTTCACAGAGTGAATCTTGAAGTCTGATGCTACATGAT 1319  
QY 1201 GGGGAGCTGTAATCAAAATGCCAAAGTAGCTGATGATGATGAGAGCTTTAAATAGAGTAGAT 1260  
DB 1320 GGGGAGCTGTAATCAAAATGCCAAAGTAGCTGATGATGATGAGAGCTTTAAATAGAGTAGAT 1379  
QY 1261 GATATTTCTGCTTCTTCAAGAGAAATAGACTTCTGCGCAGATATCTCATGAGGCTTAA 1320  
DB 1380 GATATTTCTGCTTCTTCAAGAGAAATAGACTTCTGCGCAGATATCTCATGAGGCTTAA 1439  
QY 1321 ATATGTAAGAGTAAGAGAGTTCATCTCAATCAGTGAAGATATATTTGAAGCAAAATA 1380  
DB 1440 ATATGTAAGAGTAAGAGAGTTCATCTCAATCAGTGAAGATATATTTGAAGCAAAATA 1499  
QY 1381 TTTGGAAAAACCTATGGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAACCTGAAAAAT 1440  
DB 1500 TTTGGAAAAACCTATGGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAACCTGAAAAAT 1559  
QY 1441 CTATATATGAGAGATTTTGTACTGAGCCAGAGATTAATACAGAGGCTCCCTCACAAAT 1500  
DB 1560 CTATATATGAGAGATTTTGTACTGAGCCAGAGATTAATACAGAGGCTCCCTCACAAAT 1619  
QY 1501 AAATTTAAGCGTAAAGAGAGCTTACATCAGGCTTCAATCTGAGATTTTATCAAGAAA 1560  
DB 1620 AAATTTAAGCGTAAAGAGAGCTTACATCAGGCTTCAATCTGAGATTTTATCAAGAAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAAGGAACTAACCAACGAGAG 1620  
DB 1680 GCAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAAGGAACTAACCAACGAGAG 1739  
QY 1621 CAGAATGTCAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680  
DB 1740 CAGAATGTCAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1799

1681 TCTATTCAGAAATGAGAAAAATCTTAACCCAAATAGAAATCACTGCAAAAAAGAAATCTGCTTC 1740  
1800 TCTATTCAGAAATGAGAAAAATCTTAACCCAAATAGAAATCACTGCAAAAAAGAAATCTGCTTC 1859  
1741 AAAAGAAAGCTGAACTTAAGCAGCAGATTAAGCAATTAAGAACTTCCGAATTAATATC 1800  
1860 AAAAGAAAGCTGAACTTAAGCAGCAGATTAAGCAATTAAGAACTTCCGAATTAATATC 1919  
1801 CCAATTTCAAAAGCACTAAAAAGATAGCTGAGAGAGAAAGCTTCTACAGCAATAT 1860  
1920 CCAATTTCAAAAGCACTAAAAAGATAGCTGAGAGAGAAAGCTTCTACAGCAATAT 1979  
1861 CATGGCTTGAAGCTAGTACTAGTAAATCTAAGCCCACTAATTTGACTGAATTCGA 1920  
1980 CATGGCTTGAAGCTAGTACTAGTAAATCTAAGCCCACTAATTTGACTGAATTCGA 2039  
1921 ATTGATAGTGTCTTACAGAGTGAAGATTAAGAAAAAGTACAAACCAATATGCTCA 1980  
2040 ATTGATAGTGTCTTACAGAGTGAAGATTAAGAAAAAGTACAAACCAATATGCTCA 2099  
1981 AGGCACAGAGAACTTACAGCTCATGAGAGTAAAGAACTGCACTGGAGCCAGAAAG 2040  
2100 AGGCACAGAGAACTTACAGCTCATGAGAGTAAAGAACTGCACTGGAGCCAGAAAG 2159  
2041 AGTAAACAAGCCAAATGAAACAGACAAAGTAAAGACATGATCTTCCAGAGCTG 2100  
2160 AGTAAACAAGCCAAATGAAACAGACAAAGTAAAGACATGATCTTCCAGAGCTG 2219  
2101 AAGTTAAACAAGTACAGCTGCTTCTTACTAAGTGTCTCAAAATCCAGTGAATTAAGAA 2160  
2220 AAGTTAAACAAGTACAGCTGCTTCTTACTAAGTGTCTCAAAATCCAGTGAATTAAGAA 2279  
2161 TTTGTCAATCTTACAGCTTCCAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGT 2220  
2280 TTTGTCAATCTTACAGCTTCCAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGT 2339  
2221 TCTAATTAATGCTGAAGAGCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCAGAACT 2280  
2340 TCTAATTAATGCTGAAGAGCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTTCAGAACT 2339  
2281 GAAATATCTGTAGAGTACAGATTTTCAATGGTACCTGATGATTAAGCACTGAG 2340  
2400 GAAATATCTGTAGAGTACAGATTTTCAATGGTACCTGATGATTAAGCACTGAG 2459  
2341 GAAATATCTGTAGAGTACAGATTTTCAATGGTACCTGATGATTAAGCACTGAG 2400  
2460 GAAATATCTGTAGAGTACAGATTTTCAATGGTACCTGATGATTAAGCACTGAG 2519  
2401 TGTGTAGTCACTGTGACAGCAATTTGAAAAACCCAGAGCACTAATTTGTTGCTCAAA 2460  
2520 TGTGTAGTCACTGTGACAGCAATTTGAAAAACCCAGAGCACTAATTTGTTGCTCAAA 2579  
2461 GATATAGAAATGACAGAGAAAGCTTAAATGATTCATGGGATGCAATTAACCCAGAT 2520  
2580 GATATAGAAATGACAGAGAAAGCTTAAATGATTCATGGGATGCAATTAACCCAGAT 2639  
2521 CGGGAACAAGCATGAAATGGAAGAAAGTGAATGATGCTCAGATTTGAGAAATACA 2580  
2640 CGGGAACAAGCATGAAATGGAAGAAAGTGAATGATGCTCAGATTTGAGAAATACA 2699  
2581 TTCAAGGTTTCAAGAGCCGATCATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2640  
2700 TTCAAGGTTTCAAGAGCCGATCATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2759  
2641 GAAATGCAACATCTGTGCCCCAGCTGAGGCTTAAAGAAACAAAGTCCAAAGGCACT 2700  
2760 GAAATGCAACATCTGTGCCCCAGCTGAGGCTTAAAGAAACAAAGTCCAAAGGCACT 2819  
2701 TTTGATGTTGAACAAAGAGAAAGAAATCAAGAGAAAGATGAGTCTAATATCAAGCCTGTA 2760  
2820 TTTGATGTTGAACAAAGAGAAAGAAATCAAGAGAAAGATGAGTCTAATATCAAGCCTGTA 2879  
2761 CAGACAGTTAATATCACTGACAGGCTTCTGTTGTTGCTCAGAAAGATTAAGCCAGTTGAT 2820

2880 CAGACAGTTAATATCACTGACAGGCTTCTGTTGTTGCTCAGAAAGATTAAGCAGTTGAT 2939  
2821 AATGCCAATATGATTCAGAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTCAAGAGC 2880  
2940 AATGCCAATATGATTCAGAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTCAAGAGC 2999  
2881 AAGCAACAGGACTCATCTCAATTAACATGAGCTTTTCAAAACCAATATGCTATA 2940  
3000 AAGCAACAGGACTCATCTCAATTAACATGAGCTTTTCAAAACCAATATGCTATA 3059  
2941 CCACACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3000  
3060 CCACACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3119  
3001 GAAACTTGTAGAAATTCATGATGCACTGCAAGAAAGAAAGGAAATGAGACATTCGA 3060  
3120 GAAACTTGTAGAAATTCATGATGCACTGCAAGAAAGAAAGGAAATGAGACATTCGA 3179  
3061 AGTACAGTGAACAATTAAGCCGTAATTAACATTAAGAAAAATGTTTAAAGAGCCAGC 3120  
3180 AGTACAGTGAACAATTAAGCCGTAATTAACATTAAGAAAAATGTTTAAAGAGCCAGC 3239  
3121 TCAAGCAATTAATTAAGTACGTTCCAGTCACTAATGAAGTGGCTCCAGTATTAATGA 3180  
3240 TCAAGCAATTAATTAAGTACGTTCCAGTCACTAATGAAGTGGCTCCAGTATTAATGA 3299  
3181 ATAGGTTCCAGTGAAGAAACATTCAGCAGACAGTAAAGTAAAGAGAGGCCAAATATG 3240  
3300 ATAGGTTCCAGTGAAGAAACATTCAGCAGACAGTAAAGTAAAGAGAGGCCAAATATG 3359  
3241 AATGCTATCTTATAGATTAAGGCTTTTGCAGCCGAGTCTATTAACAAAGTCTTCTGGA 3300  
3360 AATGCTATCTTATAGATTAAGGCTTTTGCAGCCGAGTCTATTAACAAAGTCTTCTGGA 3419  
3301 AGTAAATTTAGCATCTCAATTAAGAAAAAGCAAGATTAAGAAAGTGTTCAGAGCTT 3360  
3420 AGTAAATTTAGCATCTCAATTAAGAAAAAGCAAGATTAAGAAAGTGTTCAGAGCTT 3479  
3361 AATACAGATTTCTTCATATCTGATTTGATTAAGTAAAGCAAGCTTGGAGAGT 3420  
3480 AATACAGATTTCTTCATATCTGATTTGATTAAGTAAAGCAAGCTTGGAGAGT 3539  
3421 CATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATTAAG 3480  
3540 CATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTAATTAAG 3599  
3481 GAAATATCTAGTTTGTCTGAAATGACATTAAGAAAGTGTCTGCTTTTACGAAAGC 3540  
3600 GAAATATCTAGTTTGTCTGAAATGACATTAAGAAAGTGTCTGCTTTTACGAAAGC 3659  
3541 GTCGAAAGAGAGCTTAAAGAGAGCTGAGGCTGAGCCCTTACACCATTAATTTGGCTCAG 3600  
3660 GTCGAAAGAGAGCTTAAAGAGAGCTGAGGCTGAGCCCTTACACCATTAATTTGGCTCAG 3719  
3601 GCTTACGAGAGAGGCGCAAGAAATTAAGATCTTCAAGAGAACTTATCTAGTGAAGAT 3660  
3720 GCTTACGAGAGAGGCGCAAGAAATTAAGATCTTCAAGAGAACTTATCTAGTGAAGAT 3779  
3661 GAAAGAGCTTCCGCTTCCAGACATTTGTTATTTGTAAGTAAACATATACCTTCTCAG 3720  
3780 GAAAGAGCTTCCGCTTCCAGACATTTGTTATTTGTAAGTAAACATATACCTTCTCAG 3839  
3721 TCTACTAGCATAGACCGTGTGCTGCGAGTGTGCTGTAAGAACACAGAGAGAAATTTA 3780  
3840 TCTACTAGCATAGACCGTGTGCTGCGAGTGTGCTGTAAGAACACAGAGAGAAATTTA 3899  
3781 TTATCATTAAGAAATAGCTTAATATGATGCAAGTAAACAGTAAATATTTGCAAGAGCATCT 3840  
3900 TTATCATTAAGAAATAGCTTAATATGATGCAAGTAAACAGTAAATATTTGCAAGAGCATCT 3959  
3841 CAGGAACATCACTTATGAGAGAAACAAATGTTCTGCTAGCTTGTCTTCCAGAGTGC 3900



Db 3960 CAGGACATCACTTACTGAGGAAACAAAATGTTCTGCTAGCTGTTGTTCTTCACAGTGC 4019  
 QY 3901 AGTAATTTGAGAGACTGACTGCAAAATACAAACCCAGAGATCCTTCTTGATGGTCT 3960  
 Db 4020 AGTAATTTGAGAGACTGACTGCAAAATACAAACCCAGAGATCCTTCTTGATGGTCT 4079  
 QY 3961 TCCAAACAAATGAGGACATCAGTCTGAAAGCCAGGAGTGGTCTGAGTGCACAAAGATTG 4020  
 Db 4080 TCCAAACAAATGAGGACATCAGTCTGAAAGCCAGGAGTGGTCTGAGTGCACAAAGATTG 4139  
 QY 4021 GTTTCAGATATGAGAAAGAGAGAGCGGCTTGGAGAGAAAATATCAAGAGAGCAAGC 4080  
 Db 4140 GTTTCAGATATGAGAAAGAGAGAGCGGCTTGGAGAGAAAATATCAAGAGAGCAAGC 4199  
 QY 4081 ATGATATCAAACTTAGGTGAGAGAGCATCTGGGGTGTGAGAGTGAACAAAGCTCTGAA 4140  
 Db 4200 ATGATATCAAACTTAGGTGAGAGAGCATCTGGGGTGTGAGAGTGAACAAAGCTCTGAA 4259  
 QY 4141 GACTGCTCAGGGGCTATCCTCTCAGAGTGCATTTTAACTCAGCAGAGAGGATACCATG 4200  
 Db 4260 GACTGCTCAGGGGCTATCCTCTCAGAGTGCATTTTAACTCAGCAGAGAGGATACCATG 4319  
 QY 4201 CAACATTAACCTGATTAAGGCTCCAGCAGGAAATGGCTAACTAGAAAGCTGTGTAGAAACAG 4260  
 Db 4320 CAACATTAACCTGATTAAGGCTCCAGCAGGAAATGGCTAACTAGAAAGCTGTGTAGAAACAG 4379  
 QY 4261 CATGGAGGAGGAGCTTCTTAACAGTACCTTCCATCATTAAGTGAAGTGAAGTGAAGTGAAG 4320  
 Db 4380 CATGGAGGAGGAGCTTCTTAACAGTACCTTCCATCATTAAGTGAAGTGAAGTGAAGTGAAG 4439  
 QY 4321 GACCTGGGAAATCCAGAAACAAAGCAGATCAGAAAAGAGAGTATTAATCTTACAGAAAGAGT 4380  
 Db 4440 GACCTGGGAAATCCAGAAACAAAGCAGATCAGAAAAGAGAGTATTAATCTTACAGAAAGAGT 4499  
 QY 4381 AGTGAATACCTTAAAGCCAGGAAATCCAGAAAGGCTTCTCTGTCAGCAAGTTGAGTGTCT 4440  
 Db 4500 AGTGAATACCTTAAAGCCAGGAAATCCAGAAAGGCTTCTCTGTCAGCAAGTTGAGTGTCT 4559  
 QY 4441 GCAGATAGTCTTACAGTAAATAAAGAACCCAGAGAGTGGAAAGGATCCCTTCTTAAA 4500  
 Db 4560 GCAGATAGTCTTACAGTAAATAAAGAACCCAGAGAGTGGAAAGGATCCCTTCTTAAA 4619  
 QY 4501 TGGCCATCATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4560  
 Db 4620 TGGCCATCATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4679  
 QY 4561 AACTACCATCTCAAGAGAGAGTCAATTAAGTGTGTATGTGAAGAGCAACAGCTGGAA 4620  
 Db 4680 AACTACCATCTCAAGAGAGAGTCAATTAAGTGTGTATGTGAAGAGCAACAGCTGGAA 4739  
 QY 4621 GAGTCTGGGCGACAGATTTGAGCGAAACATTTACTTGGCAAGGCAAGATCTAGAGGGA 4680  
 Db 4740 GAGTCTGGGCGACAGATTTGAGCGAAACATTTACTTGGCAAGGCAAGATCTAGAGGGA 4799  
 QY 4681 ACCCTTACCTGGAATCTGGAATCAGCCTTCTCTGATGACCCCTGAATCTGATCTCT 4740  
 Db 4800 ACCCTTACCTGGAATCTGGAATCAGCCTTCTCTGATGACCCCTGAATCTGATCTCT 4859  
 QY 4741 GAAGACAGAGCCCAAGAGTCACTCGTGTGGCAACATTAACATCTTAACCTTGCAATTG 4800  
 Db 4860 GAAGACAGAGCCCAAGAGTCACTCGTGTGGCAACATTAACATCTTAACCTTGCAATTG 4919  
 QY 4801 AAAGTTCCTCCCAATTTGAAGTGTGCAAGATCTGCCAGAGTCCAGTGTCTGCTACTACT 4860  
 Db 4920 AAAGTTCCTCCCAATTTGAAGTGTGCAAGATCTGCCAGAGTCCAGTGTCTGCTACTACT 4979  
 QY 4861 GATAGTCTGGTGTATATGCAATGGAAGAGTGTGAGCAGAGGAGAGAGCCAGAAATTGACA 4920  
 Db 4980 GATAGTCTGGTGTATATGCAATGGAAGAGTGTGAGCAGAGGAGAGCCAGAAATTGACA 5039  
 QY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATCTCATGTGTGTGTGCTGACCCAGAA 4980  
 Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATCTCATGTGTGTGTGCTGACCCAGAA 5099

QY 4981 GAATTTATGCTCGTGTACAAAGTTTGCCAGAAACACCATCACTTAACTAATCTAATT 5040  
 Db 5100 GAATTTATGCTCGTGTACAAAGTTTGCCAGAAACACCATCACTTAACTAATCTAATT 5159  
 QY 5041 ACTGAGAGACTACTCATGTTGTTATGAAAACAGATCTGAGTTGTGTGTAAGGACA 5100  
 Db 5160 ACTGAGAGACTACTCATGTTGTTATGAAAACAGATCTGAGTTGTGTGTAAGGACA 5219  
 QY 5101 CTGAAATATTTTCTAGCAATTTGCCGGAGGAAAATGGTAGTTAGCTATTTCTGGGTACC 5160  
 Db 5220 CTGAAATATTTTCTAGCAATTTGCCGGAGGAAAATGGTAGTTAGCTATTTCTGGGTACC 5279  
 QY 5161 CAGTCTATTTAAAGAAAGAAAATGCTTAATGAGCATGATTTTGAAGTCAGAGAGATGTG 5220  
 Db 5280 CAGTCTATTTAAAGAAAGAAAATGCTTAATGAGCATGATTTTGAAGTCAGAGAGATGTG 5339  
 QY 5221 GTCAATGGAAGAAACCAAGAGTCCAAAGCGAGCAAGAGAAATCCAGAGCAAGAAATC 5280  
 Db 5340 GTCAATGGAAGAAACCAAGAGTCCAAAGCGAGCAAGAGAAATCCAGAGCAAGAAATC 5399  
 QY 5281 TTCAAGGGGCTAGAAATCTGTGCTATGGGCCCTTTCACCAATGCCCCACAGATCACTG 5340  
 Db 5400 TTCAAGGGGCTAGAAATCTGTGCTATGGGCCCTTTCACCAATGCCCCACAGATCACTG 5459  
 QY 5341 GAATGATGATGACAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5400  
 Db 5460 GAATGATGATGACAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5519  
 QY 5401 GGCACAGGTGTCCACCAATTTGTTGTGTGTCAGCCAGATGCTTGCAGAGAGCAATGGC 5460  
 Db 5520 GGCACAGGTGTCCACCAATTTGTTGTGTGTCAGCCAGATGCTTGCAGAGAGCAATGGC 5579  
 QY 5461 TTCCATGCAATTTGGGCGAGATGTGTGAGGCACCTGTGTGTGAGCCGAGAGTGGTGTGAC 5520  
 Db 5580 TTCCATGCAATTTGGGCGAGATGTGTGAGGCACCTGTGTGTGAGCCGAGAGTGGTGTGAC 5639  
 QY 5521 AGTGTACACTCTCAAGTGCAGAGAGTGGAGACCTGACCTGATGATGATGATGATGATGATG 5580  
 Db 5640 AGTGTACACTCTCAAGTGCAGAGAGTGGAGACCTGACCTGATGATGATGATGATGATGATG 5699  
 QY 5581 AGCCACTAC 5589  
 Db 5700 AGCCACTAC 5708

RESULT 3  
 AAT17493  
 ID AAT17493 standard; cDNA; 5914' BP.  
 AC AAT17493:  
 XX  
 AC AAT17493:  
 DT 02-OCT-1996 (first entry)  
 XX  
 DE Mutated BRCA1 coding sequence from PM06.  
 XX  
 KW Cancer therapy: breast and ovarian cancer predisposing gene; immunogen;  
 KW antibody production; germline alteration; probe; lesion neoplasia; human;  
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.  
 OS Homo sapiens.  
 OS  
 FH Key location/Qualifiers.  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= BRCA1 mutant  
 FT mutation 2731  
 FT /\*tag= b  
 FT /note= "C to T mutation"  
 PN  
 XX  
 PD W09605306-A2.  
 22-FEB-1996.

xx 11-AUG-1995; 95WO-US10202.  
xx  
xx 07-JUN-1995; 95US-0483553.  
xx 12-AUG-1994; 94US-0289221.  
xx 02-SEP-1994; 94US-0300266.  
xx 16-SEP-1994; 94US-0308104.  
xx 29-NOV-1994; 94US-0348824.  
xx 24-MAR-1995; 95US-0409305.  
xx 07-JUN-1995; 95US-0480784.  
xx  
xx (MRI-) MRIAD GENETICS INC.  
xx (CANC-) CANCER INST.  
xx (RECH-) CENT RECH DU CHUL.  
xx  
xx Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;  
xx WPI. 1996-139702/14.  
xx P-PSDB; AAR81535.  
xx  
xx New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
xx pt gene - for diagnosis and therapy of human breast and ovarian cancer  
xx and for diagnosing pre-disposition to these cancers  
xx  
xx Claim 1; 218pp; English.  
xx  
xx AAT17439-T17453 and AAT17455-T17529 represent mutations of the human  
xx breast and ovarian cancer predisposing gene (BRCA1) (for wild type see  
xx AAT17438). Proteins encoded by these mutations (see AAR81483-R81497 and  
xx AAR81499-R81546) can be used as immunogens for antibody production.  
xx These mutant BRCA1 genes have at least 1 mutation or polymorphism in  
xx comparison to the wild type sequence. By detecting a germline alteration  
xx in the wild type BRCA1 gene, a predisposition for breast and ovarian  
xx cancer can be diagnosed. In one method, BRCA1 mRNA isolated from a tissue  
xx sample from a subject has a probe, corresponding to a fragment of the  
xx wild type sequence (or an allele-specific probe for one of these  
xx mutations), added to it. The conditions allow for hybridisation of the  
xx probe to the mRNA, and any hybridisation which occurs is detected.  
xx Alternatively the BRCA1 gene in the tissue sample is isolated, and a  
xx shift in electrophoretic mobility of single stranded DNA from the sample  
xx on a non-denaturing polyacrylamide gel indicates a mutation. These  
xx methods of detection can also diagnose a lesion neoplasia associated with  
xx the BRCA1 locus. The methods may be used in gene therapy, protein  
xx replacement therapy and protein mimetics, and may be used to screen for  
xx drugs in cancer therapy.  
xx  
xx Sequence 5914 BP; 2006 A; 1155 C; 1316 G; 1437 T; 0 other:  
SQ  
Query Match 100.0%; Score 5587.4; DB 17; Length 5914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 ATGATTTATCTGCTCTTCCGCTGGAAGATACAAATGCTATTATGCTATGCGAATA 60  
DB 120 ATGGATTTATCTGCTCTTCCGCTGGAAGATACAAATGCTATTATGCTATGCGAATA 179  
OY 61 ATCTTAGAGTGTCCATCTGTCTGTGAGTGCATCAGAACCTGTCTCCACAAGTGTGAC 120  
DB 180 ATCTTAGAGTGTCCATCTGTCTGTGAGTGCATCAGAACCTGTCTCCACAAGTGTGAC 239  
OY 121 CACATATTTTGCATAATTTTGCATGCTGAAACTTCTCACCCAGAGAAAGGCGCTTCACAG 180  
DB 240 CACATATTTTGCATAATTTTGCATGCTGAAACTTCTCACCCAGAGAAAGGCGCTTCACAG 299  
OY 181 TGCTCTTATATTAAGATATATTAACCAAGAGAGCTCAAGAAAGTACGATATTAGT 240  
DB 300 TGCTCTTATATTAAGATATATTAACCAAGAGAGCTCAAGAAAGTACGATATTAGT 359  
OY 241 CAACCTGTGAAGAGCTATTTGAAATCATTTTGTCTTACAGTTGACACAGGTTTGGAG 300  
DB 360 CAACCTGTGAAGAGCTATTTGAAATCATTTTGTCTTACAGTTGACACAGGTTTGGAG 419  
OY 301 TATGCAAAACGCTATTAATTTTGCACAAAAGAAATTAAGTCTTCCGACATCTAAAGAT 360

DB 420 TATGCAAAACGCTATTAATTTTGCACAAAAGAAATTAAGTCTTCCGACATCTAAAGAT 479  
OY 361 GAAGTTTCTATCATCCAAAGATATGGCTACAGAAACCGTCCCAAAAGACTTTACAGAT 420  
DB 480 GAAGTTTCTATCATCCAAAGATATGGCTACAGAAACCGTCCCAAAAGACTTTACAGAT 539  
OY 421 GAACCCGAAATATCTCTTGTGAGAAACAGTCTAGTCTCAACTCTCAACTTGA 480  
DB 540 GAACCCGAAATATCTCTTGTGAGAAACAGTCTAGTCTCAACTCTCAACTTGA 599  
OY 481 ACTGTGAGAACTCTGAGCAAAAGCAGCGATACAACTCAAAAGAGTCTGTACATT 540  
DB 600 ACTGTGAGAACTCTGAGCAAAAGCAGCGATACAACTCAAAAGAGTCTGTACATT 659  
OY 541 GAATGGAGATCTGATCTCTCTGAGATACCGTTAATTAAGCAACTTATGCGAGTGGGA 600  
DB 660 GAATGGAGATCTGATCTCTCTGAGATACCGTTAATTAAGCAACTTATGCGAGTGGGA 719  
OY 601 GATCAAGAAATTTGTACAAATCACCCCTCAAGAGACCGAGGATGAATCAGTTGGATTCT 660  
DB 720 GATCAAGAAATTTGTACAAATCACCCCTCAAGAGACCGAGGATGAATCAGTTGGATTCT 779  
OY 661 GCAAAAAGGCTGCTGTGATTTTCTGAGACGAGATGTACAAATCTGAACATCATCA 720  
DB 780 GCAAAAAGGCTGCTGTGATTTTCTGAGACGAGATGTACAAATCTGAACATCATCA 839  
OY 721 CCCAGTAATTAATTTGAAACCCACTGGAAGCCGAGCTGAGAGGATCCACAAAAG 780  
DB 840 CCCAGTAATTAATTTGAAACCCACTGGAAGCCGAGCTGAGAGGATCCACAAAAG 899  
OY 781 TATCAGGATAGTCTGTTCAAACTGTGATGAGCCATGAGGCAAAATCTCATGCC 840  
DB 900 TATCAGGATAGTCTGTTCAAACTGTGATGAGCCATGAGGCAAAATCTCATGCC 959  
OY 841 AGCTCATTAACAGCATGAGAACAGCAGTTTATTACTACTAAGACAGATGANTAGAA 900  
DB 960 AGCTCATTAACAGCATGAGAACAGCAGTTTATTACTACTAAGACAGATGANTAGAA 1019  
OY 901 AAGCGTGAATTCGTATTAATAAGCAAAACAGCCGTGTTGCAAGAGCCACATTAACGA 960  
DB 1020 AAGCGTGAATTCGTATTAATAAGCAAAACAGCCGTGTTGCAAGAGCCACATTAACGA 1079  
OY 961 TGGGCTGGAATGAAGAAACATGTAATGATAGCGGACTCCAGCAGCAAGAAAAAGSTA 1020  
DB 1080 TGGGCTGGAATGAAGAAACATGTAATGATAGCGGACTCCAGCAGCAAGAAAAAGSTA 1139  
OY 1021 GATCTGAATTCGTATTCCTCTGTGTGAGAGAAAGATGGAATTAAGCAGAACTGCATGC 1080  
DB 1140 GATCTGAATTCGTATTCCTCTGTGTGAGAGAAAGATGGAATTAAGCAGAACTGCATGC 1199  
OY 1081 TCAGAGAAATCTAGAGATCTAGAGATGTTCCCTGTGATTAACCTAATTAAGCAGTTGAG 1140  
DB 1200 TCAGAGAAATCTAGAGATCTAGAGATGTTCCCTGTGATTAACCTAATTAAGCAGTTGAG 1259  
OY 1141 AAGGTTAATGAGTGTGTTTCCAGAACTGATGAGTGTAGTTCGTGATGACTCACATGAT 1200  
DB 1260 AAGGTTAATGAGTGTGTTTCCAGAACTGATGAGTGTAGTTCGTGATGACTCACATGAT 1319  
OY 1201 GGGGAGTCTGAATCAATATCCCAAAGTACGTGATTTGAGAGCTTTAAAGTACAGTAT 1260  
DB 1320 GGGGAGTCTGAATCAATATCCCAAAGTACGTGATTTGAGAGCTTTAAAGTACAGTAT 1379  
OY 1261 GAATATTCGTGTTCTCAGAGAAATATAGCTACAGGCGAGGATCCCATGAGGCTTTA 1320  
DB 1380 GAATATTCGTGTTCTCAGAGAAATATAGCTACAGGCGAGGATCCCATGAGGCTTTA 1439  
OY 1321 ATATGTAAGAGTGAAGAGTCTCACTCCAAATCAGTAGAGAGTAAATATTGAAGACAAATA 1380  
DB 1440 ATATGTAAGAGTGAAGAGTCTCACTCCAAATCAGTAGAGAGTAAATATTGAAGACAAATA 1439  
OY 1381 TTTGGGAAACCTATCAGAGAGAGGAGCCCTCCCACTTAAGCATGTAACTGAATAT 1440

1500 TTGGGAAAACCTATCGAGAGAAGCAAGCCTCCCAACTTAAGCATATGTAAGTAAT 1559  
1441 CTAATTATAGAGAGCTTTGTTACTGAGCCACAGATTAATACAGAGGTCCTCCACAAAT 1500  
1560 CTAAATTATAGAGCAATTTGTTACTGAGCCACAGATTAATACAGAGGTCCTCCACAAAT 1619  
1501 AAATTAAAGCGTAAAGAGAGACCTACATAGAGCCTTCATCTGAGAGTTTATCAAGAA 1560  
1620 AATTAAAGCGTAAAGAGAGACCTACATAGAGCCTTCATCTGAGAGTTTATCAAGAA 1679  
1561 GCAGATTTGGCAGTTCACAAAGAGCTCCTGAATGATTAATACAGAGAACTACCAAAACGAG 1620  
1680 GCAGATTTGGCAGTTCACAAAGAGCTCCTGAATGATTAATACAGAGAACTACCAAAACGAG 1739  
1621 CAGAAATGTCAGTGTGAATTAATTAATAGTGCATGAGAAATTAACAAAGGTGAT 1680  
1740 CAGAAATGTCAGTGTGAATTAATTAATAGTGCATGAGAAATTAACAAAGGTGAT 1799  
1681 TCTATTGAGAAATGAGAAAATCTTAACCAATAGAAATCACTGCAAAAAGATCTGCTTC 1740  
1800 TCTATTGAGAAATGAGAAAATCTTAACCAATAGAAATCACTGCAAAAAGATCTGCTTC 1859  
1741 AAAAGCAAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAAGTCAATTAATATC 1800  
1860 AAAAGCAAAAGCTGAACCTTAAGCAGCAGTATAGCAATATGGAAGTCAATTAATATC 1919  
1801 CACAATTCAAAAGCCTTAAGAAATAGGCTGAGAGAGAGTCTTCCACGAGCATAT 1860  
1920 CACAATTCAAAAGCCTTAAGAAATAGGCTGAGAGAGAGTCTTCCACGAGCATAT 1979  
1861 CATGCGCTTGAACCTAGTACTAGTAAATCTAAGCCCACTAATTTGTAATGCAA 1920  
1980 CATGCGCTTGAACCTAGTACTAGTAAATCTAAGCCCACTAATTTGTAATGCAA 2039  
1921 ATTGATATGTTGTTCTAGCAGTAAGATTAAGAAAAAAGTACCAACCAATGCCAGTC 1980  
2040 ATTGATATGTTGTTCTAGCAGTAAGATTAAGAAAAAAGTACCAACCAATGCCAGTC 2099  
1981 AAGCAGAGAGAAACCTCAACCTAGAAAGTAAAGAACTGCAACTGAGCCAGCAAG 2040  
2100 AAGCAGAGAGAAACCTCAACCTAGAAAGTAAAGAACTGCAACTGAGCCAGCAAG 2159  
2041 AATAACCAAGCAAAATGAAAGAGAGAAAGTAAAGACATGAGTGAATCTTCCAGAGCTG 2100  
2160 AATAACCAAGCAAAATGAAAGAGAGAAAGTAAAGACATGAGTGAATCTTCCAGAGCTG 2219  
2101 AAGTTAACAATAGCAGCTGCTGTTCTTTTACTAAGTGTCAATATCCAGTAACTAAAGAA 2160  
2220 AAGTTAACAATAGCAGCTGCTGTTCTTTTACTAAGTGTCAATATCCAGTAACTAAAGAA 2279  
2161 TTGTGCAATCTAGGCTTCCAGAGAGAAAGAAAGAACTGAAACAGTTAAAGTG 2220  
2280 TTGTGCAATCTAGGCTTCCAGAGAGAAAGAAAGAAAGTAAAGTAAAGTG 2339  
2221 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAAACT 2280  
2340 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTTGCAAACT 2399  
2281 GAAAGATCTGTAGAGAGTGAAGATTTTATGTTGTAAGTCTGTAAGTATATGCACTGAG 2340  
2400 GAAAGATCTGTAGAGAGTGAAGATTTTATGTTGTAAGTCTGTAAGTATATGCACTGAG 2459  
2341 GAAAGATCTGTAGAGAGTGAAGTATGCACTGTAAGGAAAGGCAAAAACAACCAATATA 2400  
2460 GAAAGATCTGTAGAGAGTGAAGTATGCACTGTAAGGAAAGGCAAAAACAACCAATATA 2519  
2401 TGTGTGAGTCACTGTGACACATTTTGAAGAAACCCCAAGGAGTAAATCATGTTGTTCCAA 2460  
2520 TGTGTGAGTCACTGTGACACATTTTGAAGAAACCCCAAGGAGTAAATCATGTTGTTCCAA 2579  
2461 GATTAATGAATGACAGAGAGCTTTAAGTATCCATTGGACATGAACTTAACCAAGT 2520  
2580 GATTAATGAATGACAGAGAGCTTTAAGTATCCATTGGACATGAACTTAACCAAGT 2639

2521 CGGGAACAAGCATAGAAATGCAAGAAAGTGAACCTTGATGCTCAGTATTTGCAGATPAC 2580  
2640 CGGGAACAAGCATAGAAATGCAAGAAAGTGAACCTTGATGCTCAGTATTTGCAGATPAC 2699  
2581 TTCAAGTTTCAAAAGCGCAGTCAATTTGCTGTTTTCAAATCCAGAGAAATGCAAGAG 2640  
2700 TTCAAGTTTCAAAAGCGCAGTCAATTTGCTGTTTTCAAATCCAGAGAAATGCAAGAG 2759  
2641 GAATGTGCAACATTTCTGCTGCCACTGCTGGTCTTAAAGAAACAAAGTCCAAAGTCACT 2700  
2760 GAATGTGCAACATTTCTGCTGCCACTGCTGGTCTTAAAGAAACAAAGTCCAAAGTCACT 2819  
2701 TTTGAATGTGAACAAAGAGAGAAATTAAGAGAAATAGAGTCAATATCAAGCCTGTA 2760  
2820 TTTGAATGTGAACAAAGAGAGAAATTAAGAGAAATAGAGTCAATATCAAGCCTGTA 2879  
2761 CAGACAGTAAATATCACTGAGCCTTCTGTTGTTGTCAGAAAGATAGCCAGTTGAT 2820  
2880 CAGACAGTAAATATCACTGAGCCTTCTGTTGTTGTCAGAAAGATAGCCAGTTGAT 2939  
2821 AATGCCAATGTATGATCAAGAGAGCTAGGTTTGTCTATCATCTCAGTTCAAGGC 2880  
2940 AATGCCAATGTATGATCAAGAGAGCTAGGTTTGTCTATCATCTCAGTTCAAGGC 2999  
2881 AAGCAAGTGGAGCTCATTTCTCCAAATTAACATGAGACTTTTACAAAACCATATGATTA 2940  
3000 AAGCAAGTGGAGCTCATTTCTCCAAATTAACATGAGACTTTTACAAAACCATATGATTA 3059  
2941 CCACCACTTTTCCATCAAGTCAATTTGTTAAACATAATGTAAAGAAATCTGCTAGAG 3000  
3060 CCACCACTTTTCCATCAAGTCAATTTGTTAAACATAATGTAAAGAAATCTGCTAGAG 3119  
3001 GAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAAATGGAAATGAGAACTTCCA 3060  
3120 GAAACTTTGAGAGACATTCATGTCACCTGAAAGAGAAATGGAAATGAGAACTTCCA 3179  
3061 AGTACAGTGGAGCAATTTAGCGGTATTAACATTAAGAGAAATGTTTAAAGAGGCAGC 3120  
3180 AGTACAGTGGAGCAATTTAGCGGTATTAACATTAAGAGAAATGTTTAAAGAGGCAGC 3239  
3121 TCAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3180  
3240 TCAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3299  
3181 ATAGGTTCCAGTATTAAGAAACATTCAGCAGAACTAGTGAAGAAAGAGGCCCCAAATTTG 3240  
3300 ATAGGTTCCAGTATTAAGAAACATTCAGCAGAACTAGTGAAGAAAGAGGCCCCAAATTTG 3359  
3241 AATGCTATGCTTATGATTAAGGAGTTTTCAGAACTGAGAGTCTATTAACAAAGTCTTCCGGA 3300  
3360 AATGCTATGCTTATGATTAAGGAGTTTTCAGAACTGAGAGTCTATTAACAAAGTCTTCCGGA 3419  
3301 AGTAATTTGAAGCATCTGAATTAAGAAAGCAAGAAATGAAGAAAGTATGACAGTGT 3360  
3420 AGTAATTTGAAGCATCTGAATTAAGAAAGCAAGAAATGAAGAAAGTATGACAGTGT 3479  
3361 AATACAGATTTCTCTCATATCTGATTTCAAGATTAACCTTGAACAGCTATGGAATAGT 3420  
3480 AATACAGATTTCTCTCATATCTGATTTCAAGATTAACCTTGAACAGCTATGGAATAGT 3539  
3421 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATATAGTGAATATAAG 3480  
3540 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATATAGTGAATATAAG 3599  
3481 GAGATATAGTATTTGCTGAAATGAACATTAAGAAAGTCTGCTTTTATAGCAAAAGC 3540  
3600 GAGATATAGTATTTGCTGAAATGAACATTAAGAAAGTCTGCTTTTATAGCAAAAGC 3659  
3541 GTCCAGAAAGAGAGGCTTGAAGAGAGTCTAGCCCTTTACCCATACACATTTGGCTCAG 3600  
3660 GTCCAGAAAGAGAGGCTTGAAGAGAGTCTAGCCCTTTACCCATACACATTTGGCTCAG 3719

QY 3601 GGTTCAGAGAGGGGCCAAGAAATTAGACTCCTCAGAGAGAACTTATCTAGTAGCAT 3660  
| | | | |  
Db 3720 GGTTCAGAGAGGGGCCAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTAGCAT 3779  
| | | | |  
QY 3661 GAAAGAGCTCCCTGCTCCAACTGTATTTGGTAAAGTAAACATATACCTTCAG 3720  
| | | | |  
Db 3780 GAAAGAGCTCCCTGCTCCAACTGTATTTGGTAAAGTAAACATATACCTTCAG 3839  
| | | | |  
QY 3721 TCTACTAGGCAATAGACCGCTTGCTACCGAGTGTCTGTACAGACAGAGAGAAATTTA 3780  
| | | | |  
Db 3840 TCTACTAGGCAATAGACCGCTTGCTACCGAGTGTCTGTACAGACAGAGAGAAATTTA 3899  
| | | | |  
QY 3781 TTATCATTTAGAAATAGCTTAAATAGCTGCAATACAGGTAAATTTGGCAAGGCACT 3840  
| | | | |  
Db 3900 TTATCATTTAGAAATAGCTTAAATAGCTGCAATACAGGTAAATTTGGCAAGGCACT 3959  
| | | | |  
QY 3841 CAGGAACATCACCCTTGCTAGGAAACAAATGTTTGTCTAGTCTGTTTCTTACAGTGC 3900  
| | | | |  
Db 3960 CAGGAACATCACCCTTGCTAGGAAACAAATGTTTGTCTAGTCTGTTTCTTACAGTGC 4019  
| | | | |  
QY 3901 ACTGAATTTGGAAGACTTGACTGCAATACAAACACCAGATCCTTCTGATTGGTTCT 3960  
| | | | |  
Db 4020 ACTGAATTTGGAAGACTTGACTGCAATACAAACACCAGATCCTTCTGATTGGTTCT 4079  
| | | | |  
QY 3961 TCCAAACAAATGAGGATCATGCTGAAAGCCAGGGAGTTGCTGAGTGACAAGAAATG 4020  
| | | | |  
Db 4080 TCCAAACAAATGAGGATCATGCTGAAAGCCAGGGAGTTGCTGAGTGACAAGAAATG 4139  
| | | | |  
QY 4021 GTTTCAGATGATGAAGAGAGAGAGGAGGCTTGAGAAATTAATCAAGAGAGCAAGC 4080  
| | | | |  
Db 4140 GTTTCAGATGATGAAGAGAGAGAGGAGGCTTGAGAAATTAATCAAGAGAGCAAGC 4199  
| | | | |  
QY 4081 ATGATTTCAACTTAGTGTAAGCAGCATCTGGGTGTGAGAGTAAACAGCCGTCTGTGA 4140  
| | | | |  
Db 4200 ATGATTTCAACTTAGTGTAAGCAGCATCTGGGTGTGAGAGTAAACAGCCGTCTGTGA 4259  
| | | | |  
QY 4141 GACTGCTGAGGGGCTATCTCTCAGAGTGCATTTTAAACCACTCAGAGAGGGATACCATG 4200  
| | | | |  
Db 4260 GACTGCTGAGGGGCTATCTCTCAGAGTGCATTTTAAACCACTCAGAGAGGGATACCATG 4319  
| | | | |  
QY 4201 CACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAAGTACAGTGTGTTAGAACAG 4260  
| | | | |  
Db 4320 CACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAAGTACAGTGTGTTAGAACAG 4379  
| | | | |  
QY 4261 CATGGAGCCAGCCCTTCTAAGCCTACCTTCATCATAGTACTCTTCTGCCCTTGAAG 4320  
| | | | |  
Db 4380 CATGGAGCCAGCCCTTCTAAGCCTACCTTCATCATAGTACTCTTCTGCCCTTGAAG 4439  
| | | | |  
QY 4321 GACCTGCGGAAATCCAGAAACAGCATCAGAAAAAGCAGTATTAACCTTCACAGAAAGT 4380  
| | | | |  
Db 4440 GACCTGCGGAAATCCAGAAACAGCATCAGAAAAAGCAGTATTAACCTTCACAGAAAGT 4499  
| | | | |  
QY 4381 AGTGAATACCTATTAAGCCAGAAATCCAGAAAGCCCTTCTGCTACAGTTTGAAGTGTCT 4440  
| | | | |  
Db 4500 AGTGAATACCTATTAAGCCAGAAATCCAGAAAGCCCTTCTGCTACAGTTTGAAGTGTCT 4559  
| | | | |  
QY 4441 GCAGATAGTTTACAGTAAAAATAAAGAACAGAGATGGAAGGTTCATCCCTTCTAAA 4500  
| | | | |  
Db 4560 GCAGATAGTTTACAGTAAAAATAAAGAACAGAGATGGAAGGTTCATCCCTTCTAAA 4619  
| | | | |  
QY 4501 TGCCCATCATTTAGATGATAGTGTGTATGACACAGTGTCTGCGAGGTCTTCAAGATAGA 4560  
| | | | |  
Db 4620 TGCCCATCATTTAGATGATAGTGTGTATGACACAGTGTCTGCGAGGTCTTCAAGATAGA 4679  
| | | | |  
QY 4561 AACTACCATCTCAAGAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGGAA 4620  
| | | | |  
Db 4680 AACTACCATCTCAAGAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGGAA 4739  
| | | | |  
QY 4621 GAGTCTGGGCCACACAGATTTGAACGAAACATCTTACTTCCAGAGCAAGATCTAGAGGA 4680  
| | | | |  
Db 4740 GAGTCTGGGCCACACAGATTTGAACGAAACATCTTACTTCCAGAGCAAGATCTAGAGGA 4799  
| | | | |  
QY 4681 ACCCTTACCTGGAAATCTGGAAATCAAGCCTTCTCTGATGATACCCTGAATCTGATCTTCT 4740  
| | | | |

Db 4800 ACCCTTACCTGGAAATCTGGAAATCAAGCCTTCTCTGATGAGACCCTGAATCTGATCTCT 4859  
| | | | |  
QY 4741 GAAAG 4800  
| | | | |  
Db 4860 GAAAG 4919  
| | | | |  
QY 4801 AAGGTTCCCAATTTGAAGTTGAGAAATCTGCCAGAGTCCAGTCTGCTCTCAATCTACT 4860  
| | | | |  
Db 4920 AAGGTTCCCAATTTGAAGTTGAGAAATCTGCCAGAGTCCAGTCTGCTCTCAATCTACT 4979  
| | | | |  
QY 4861 GATAGTCTGGGTATATATGCAATGGAAGAAAGTGTGAGCAGGAGAGAGAGAGAGAGAGAG 4920  
| | | | |  
Db 4980 GATAGTCTGGGTATATATGCAATGGAAGAAAGTGTGAGCAGGAGAGAGAGAGAGAGAGAG 5039  
| | | | |  
QY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATAGTGTGTGTGCTGAGCCGACAGAA 4980  
| | | | |  
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATAGTGTGTGTGCTGAGCCGACAGAA 5099  
| | | | |  
QY 4981 GAATTTATGCTGCTGATCAAGATTGTCAGAAACACACATCATCTTAACTAATTAAT 5040  
| | | | |  
Db 5100 GAATTTATGCTGCTGATCAAGATTGTCAGAAACACACATCATCTTAACTAATTAAT 5159  
| | | | |  
QY 5041 ACTGAAGAGACTATCTATGTTGTTATGAACAGATGCTGAATTTGTGTGACGAGCA 5100  
| | | | |  
Db 5160 ACTGAAGAGACTATCTATGTTGTTATGAACAGATGCTGAATTTGTGTGACGAGCA 5219  
| | | | |  
QY 5101 CTGAATTAATTTCTAGCAATGCGGAGAGAAATGGGTGATTAATTTCTGGGTGAGC 5160  
| | | | |  
Db 5220 CTGAATTAATTTCTAGCAATGCGGAGAGAAATGGGTGATTAATTTCTGGGTGAGC 5279  
| | | | |  
QY 5161 CAGCTATTTAAAGAAAGAAATGCTGAATGAGATGATTTTGAAGTACAGAGAGATG 5220  
| | | | |  
Db 5280 CAGCTATTTAAAGAAAGAAATGCTGAATGAGATGATTTTGAAGTACAGAGAGATG 5339  
| | | | |  
QY 5221 GTCAATGGAAGAAACCCAGAGTCCAAAGCCAGAGCAAGAGATCCAGAGAGAAATG 5280  
| | | | |  
Db 5340 GTCAATGGAAGAAACCCAGAGTCCAAAGCCAGAGCAAGAGATCCAGAGAGAAATG 5399  
| | | | |  
QY 5281 TTGAGGGGCTAGAAATGCTTGTGTTGAGGCTTCCACCAATGCCCCAGATCACTG 5340  
| | | | |  
Db 5400 TTGAGGGGCTAGAAATGCTTGTGTTGAGGCTTCCACCAATGCCCCAGATCACTG 5459  
| | | | |  
QY 5341 GAATGATGGTACAGCTGTGTGTGCTTCTGTTGTTGAGAGAGCTTTCATTCACCTT 5400  
| | | | |  
Db 5460 GAATGATGGTACAGCTGTGTGTGCTTCTGTTGTTGAGAGAGCTTTCATTCACCTT 5519  
| | | | |  
QY 5401 GGCACAGGTGTCCACCAATTTGTTGTTGTCAGCCAGATGCTGTGACAGAGAGACATG 5460  
| | | | |  
Db 5520 GGCACAGGTGTCCACCAATTTGTTGTTGTCAGCCAGATGCTGTGACAGAGAGACATG 5579  
| | | | |  
QY 5461 TTTCATGCAATTTGGGAGATGTTGTCAGGACCTGTGTGTCAGCCAGAGAGAGAGAG 5520  
| | | | |  
Db 5580 TTTCATGCAATTTGGGAGATGTTGTCAGGACCTGTGTGTCAGCCAGAGAGAGAG 5639  
| | | | |  
QY 5521 AGTGAAGACTATACAGAGTCCAGAGAGCTGACACATCTGATATACCCAGATCCCCAC 5580  
| | | | |  
Db 5640 AGTGAAGACTATACAGAGTCCAGAGAGCTGACACATCTGATATACCCAGATCCCCAC 5699  
| | | | |  
QY 5581 AGCCACTAC 5589  
| | | | |  
Db 5700 AGCCACTAC 5708  
| | | | |  
RESULT 4  
AAT17495  
ID AAT17495 standard; cDNA: 5914 BP.  
XX  
AC AAT17495;  
XX  
DT 02-OCT-1996 (first entry)  
XX  
DE Mutated BRCA1 coding sequence from PM09.

XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;  
 KW antibody production; germline alteration; probe; lesion neoplasia; human;  
 KM gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= BRCA1 mutant  
 FT mutation 2201  
 FT /\*tag= b  
 FT /note= "C to T mutation"  
 XX  
 PD WO9605306-A2.  
 XX 22-FEB-1996.  
 XX 11-AUG-1995; 95MO-US10202.  
 XX  
 PR 07-JUN-1995; 95US-0483553.  
 PR 12-AUG-1994; 94US-0289221.  
 PR 02-SEP-1994; 94US-0300266.  
 PR 16-SEP-1994; 94US-0308104.  
 PR 29-NOV-1994; 94US-0348824.  
 PR 24-MAR-1995; 95US-0409305.  
 PR 07-JUN-1995; 95US-0480784.  
 XX  
 PA (MYRI-) MYRIAD GENETICS INC.  
 PA (CANC-) CANCER INST.  
 PA (RECH-) CENT RECH DU CHUL.  
 XX  
 PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;  
 DR WPI; 1996-139702/14.  
 DR P-SDB; AAR81481.  
 XX  
 PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1  
 PT gene - for diagnosis and therapy of human breast and ovarian cancer  
 PT and for diagnosing pre-disposition to these cancers  
 PS Claim 1; 218bp; English.  
 XX  
 CC AAT17439-T17453 and AAT17455-T17529 represent mutations of the human  
 CC breast and ovarian cancer predisposing gene (BRCA1) (for wild type see  
 CC AAT17438). Proteins encoded by these mutations (see AAR81483-R81497 and  
 CC AAR81499-R81546) can be used as immunogens for antibody production.  
 CC These mutant BRCA1 genes have at least 1 mutation or polymorphism in  
 CC comparison to the wild type sequence. By detecting a germline alteration  
 CC in the wild type BRCA1 gene, a predisposition for breast and ovarian  
 CC cancer can be diagnosed. In one method, BRCA1 mRNA isolated from a tissue  
 CC sample from a subject has a probe, corresponding to a fragment of the  
 CC wild type sequence (or an allele-specific probe for one of these  
 CC mutations), added to it. The conditions allow for hybridisation of the  
 CC probe to the mRNA, and any hybridisation which occurs is detected.  
 CC Alternatively the BRCA1 gene in the tissue sample is isolated, and a  
 CC shift in electrophoretic mobility of single stranded DNA from the sample  
 CC on a non-denaturing polyacrylamide gel indicates a mutation. These  
 CC methods of detection can also diagnose a lesion neoplasia associated with  
 CC the BRCA1 locus. The methods may be used in gene therapy, protein  
 CC replacement therapy and protein mimetics, and may be used to screen for  
 CC drugs in cancer therapy.  
 XX  
 SO Sequence 5914 BP; 2006 A; 1155 C; 1316 G; 1437 T; 0 other;

QY 61 ATCTTAGAGTGTCCCATCTGCTGAGTGTGATCAAGGAACCTGTCTCCAAAGTGTAC 120  
 |||||||  
 Db 180 ATCTTAGAGTGTCCCATCTGCTGAGTGTGATCAAGGAACCTGTCTCCAAAGTGTAC 239  
 QY 121 CACATATTTTGGCAAAATTTTGCATGCTGAACCTTCTCAACGAGAAAGGCTTCACAG 180  
 |||||||  
 Db 240 CACATATTTTGGCAAAATTTTGCATGCTGAACCTTCTCAACGAGAAAGGCTTCACAG 269  
 QY 181 TGTCTTTATGTATAGATGATATACCAAAAGAGCTTACAAAGAAATAGAGATTAGT 240  
 |||||||  
 Db 300 TGTCTTTATGTATAGATGATATACCAAAAGAGCTTACAAAGAAATAGAGATTAGT 359  
 QY 241 CAACCTGTGAAGAGCTATTTGAAATCATTTTGTCTTACGCTTGACAGGTTTGAG 300  
 |||||||  
 Db 360 CAACCTGTGAAGAGCTATTTGAAATCATTTTGTCTTACGCTTGACAGGTTTGAG 419  
 QY 301 TATGCAAAACAGCTATATATTTTGGCAAAAGAAATTAACCTCTGTAACATCTAAAGAT 360  
 |||||||  
 Db 420 TATGCAAAACAGCTATATATTTTGGCAAAAGAAATTAACCTCTGTAACATCTAAAGAT 479  
 QY 361 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTCCAAAGACTTCTACAGAT 420  
 |||||||  
 Db 480 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTCCAAAGACTTCTACAGAT 539  
 QY 421 GAACCCGAAATCTCTCTGCGAAGAAACAGTCTGATGCTCAACTCTTAACCTTGA 480  
 |||||||  
 Db 540 GAACCCGAAATCTCTCTGCGAAGAAACAGTCTGATGCTCAACTCTTAACCTTGA 599  
 QY 481 ACTGTGAGAACTCTGAGGACAAAGCAGGATCAACCTCAAAAGAGCTGTCTACATT 540  
 |||||||  
 Db 600 ACTGTGAGAACTCTGAGGACAAAGCAGGATCAACCTCAAAAGAGCTGTCTACATT 659  
 QY 541 GAATTGGGATCTGATTTCTTGAAGATACCGCTTAATTAAGCACTTATTGCACTGTGGA 600  
 |||||||  
 Db 660 GAATTGGGATCTGATTTCTTGAAGATACCGCTTAATTAAGCACTTATTGCACTGTGGA 719  
 QY 601 GATCAGAAATTTGTACAAATCACCCTCAAGGAAACCGGATGAATCAGTTGGATTCT 660  
 |||||||  
 Db 720 GATCAGAAATTTGTACAAATCACCCTCAAGGAAACCGGATGAATCAGTTGGATTCT 779  
 QY 661 GCAAAAAAGGCTCTGTGTAATTTTCTGAGAGGATGTAACAATATCATCAATCAA 720  
 |||||||  
 Db 780 GCAAAAAAGGCTCTGTGTAATTTTCTGAGAGGATGTAACAATATCATCAATCAA 839  
 QY 721 CCCAGTATATATTTTGAACACCACTGAGAGAGCGTGCACCTGAGAGGATCCAGAAAG 780  
 |||||||  
 Db 840 CCCAGTATATATTTTGAACACCACTGAGAGAGCGTGCACCTGAGAGGATCCAGAAAG 899  
 QY 781 TATCAGGATGTTCTGTTTCAAACTTCATGTGAGGCGATGTGACCAATATCATGCC 840  
 |||||||  
 Db 900 TATCAGGATGTTCTGTTTCAAACTTCATGTGAGGCGATGTGACCAATATCATGCC 959  
 QY 841 AGCTCATTCAGAGTGAAGAACACAGTTTATCTCACTAAAGACAGATGATAGAA 900  
 |||||||  
 Db 960 AGCTCATTCAGAGTGAAGAACACAGTTTATCTCACTAAAGACAGATGATAGAA 1019  
 QY 901 AAGCGTGAATCTGTATATAAAGCAACAGCGCTGTACAGAGGCCAATATACAGA 960  
 |||||||  
 Db 1020 AAGCGTGAATCTGTATATAAAGCAACAGCGCTGTACAGAGGCCAATATACAGA 1079  
 QY 961 TGGGCTGGAAGTGAAGAAACATGTATGATGAGCGGACCTCCAGACAGAAAAAGGTA 1020  
 |||||||  
 Db 1080 TGGGCTGGAAGTGAAGAAACATGTATGATGAGCGGACCTCCAGACAGAAAAAGGTA 1139  
 QY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATGAGCAAGATGCCATGC 1080  
 |||||||  
 Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATGAGCAAGATGCCATGC 1199  
 QY 1081 TCAGAGAAATCTGAGATGATGAGATGTTCTTGGATTAACATAATATACAGATTGAG 1140  
 |||||||  
 Db 1200 TCAGAGAAATCTGAGATGATGAGATGTTCTTGGATTAACATAATATACAGATTGAG 1259

QY	1141	AAAGTTAATGAGTGGTTTTTCCAGAAGTATGACTGTTAGTTCTGATGACTCACATGAT	1200
Db	1260	AAAGTTAATGAGTGGTTTTTCCAGAAGTATGAAAGCTTAGGTTCTGAGACTCACATGAT	1319
QY	1201	GGGGAGTCTGAATCAAAATGCCAAATGACTGATGTATTTGGAGCGTTCTTAAATGAGTGAAT	1260
Db	1320	GGGGAGTCTGAATCAAAATGCCAAATGACTGATGTATTTGGAGCGTTCTTAAATGAGTGAAT	1379
QY	1261	GAATATTCGTGTTCTTCAGAGAAATATAGACTTACTGGCCAGTGTACTCATGAGGCTTTA	1320
Db	1380	GAATATTCGTGTTCTTCAGAGAAATATAGACTTACTGGCCAGTGTACTCATGAGGCTTTA	1439
QY	1321	ATATGTAAAGTGAAGAGGTTCTCTCCAAATCTAGAGAGTAAATATTGGAAGACAAATA	1380
Db	1440	ATATGTAAAGTGAAGAGGTTCTCTCCAAATCTAGAGAGTAAATATTGGAAGACAAATA	1499
QY	1381	TTTGGGAAAACCTATTCGGAGAGAGGCAAGCCCTCCCACTTAAGCCATGTAACTGAAAAT	1440
Db	1500	TTTGGGAAAACCTATTCGGAGAGAGGCAAGCCCTCCCACTTAAGCCATGTAACTGAAAAT	1559
QY	1441	CTAATTTATAGAGCAATTTGTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAAT	1500
Db	1560	CTAATTTATAGAGCAATTTGTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAAT	1619
QY	1501	AAATTTAAAGCGTTAAAGAGACCTTCATCAGCGCTTCATCCTGAGAGATTTTATCAAGAA	1560
Db	1620	AAATTTAAAGCGTTAAAGAGACCTTCATCAGCGCTTCATCCTGAGAGATTTTATCAAGAA	1679
QY	1561	GCAGATTTGGCAGGTTCAAAAGACCTCTCAAAATGATTAATCAAGGGAACTTAACCAACGGAG	1620
Db	1680	GCAGATTTGGCAGGTTCAAAAGACCTCTCTAAATGATTAATCAAGGGAACTTAACCAACGGAG	1739
QY	1621	CAGAAATGTCAGATGATGAATATTACTAATAGTGGTCATGGAATTAACCAAGAGTGAT	1680
Db	1740	CAGAAATGTCAGATGATGAATATTACTAATAGTGGTCATGGAATTAACCAAGAGTGAT	1799
QY	1681	TCTATTTCGAATGAGAAAAATTCCTAACCCATAGAACTACTCGAAAAAGATCTGCTTTC	1740
Db	1800	TCTATTTCGAATGAGAAAAATTCCTAACCCATAGAACTACTCGAAAAAGATCTGCTTTC	1859
QY	1741	AAAACGAAGCTGAACCTATTAAGCGCGCTTAAGCAATATGGAACCTGAAATTAATATTC	1800
Db	1860	AAAACGAAGCTGAACCTATTAAGCGCGCTTAAGCAATATGGAACCTGAAATTAATATTC	1919
QY	1801	CACAATTCMAAAGCACCTTAAAGAGAAATAGGCTGAGAGAGAGTCTTACAGGACATAT	1860
Db	1920	CACAATTCMAAAGCACCTTAAAGAGAAATAGGCTGAGAGAGAGTCTTACAGGACATAT	1979
QY	1861	CATGCGCTTGAACTAGTAGTAGTAAGAAATCTAAGCCCACTAATTTGACTGAATTCAA	1920
Db	1980	CATGCGCTTGAACTAGTAGTAGTAAGAAATCTAAGCCCACTAATTTGACTGAATTCAA	2039
QY	1921	ATTGATATGTTTCTACAGAGTGAAGAGATTAAGAAAAAAGTACAAACCAATTCGCACTC	1980
Db	2040	ATTGATATGTTTCTACAGAGTGAAGAGATTAAGAAAAAAGTACAAACCAATTCGCACTC	2099
QY	1981	AGGACAGACAAACCTTACCACTCATGGAAGTAAAGAACTCGCACTGGAGCCCAAGAG	2040
Db	2100	AGGACAGACAAACCTTACCACTCATGGAAGTAAAGAACTCGCACTGGAGCCCAAGAG	2159
QY	2041	AGTAAACAAGCCAAATGAAACAGACAAGTAAAGACATGACAGTGTACTTTCCAGAGCTG	2100
Db	2160	AGTAAACAAGCCAAATGAAACAGACAAGTAAAGACATGACAGTGTACTTTCCAGAGCTG	2219
QY	2101	AAAGTTAACAAATGCACTCGTCTTTTAACTAAGTGTTCAAATACCAGTGAACCTTAAAGAA	2160
Db	2220	AAAGTTAACAAATGCACTCGTCTTTTAACTAAGTGTTCAAATACCAGTGAACCTTAAAGAA	2279
QY	2161	TTTGTGCATCTCTACGCTTCCAAAGAGAAAGAAAGAAAGAAACCTAAGAAACAGTAAAGTG	2220
Db	2280	TTTGTGCATCTCTCTACGCTTCCAAAGAGAAAGAAAGAAAGAAACCTAAGAAACAGTAAAGTG	2339
QY	2221	TCTATATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGGAGAAAGGTTTTCGAACACT	2280

Db	2340	TCATAATAAGCGTGAAGACCCCAAAATCCATGTTAAGTGGAAAGGGTTTGGCAACT	2399
OY	2281	GAAGATCTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2340
Db	2400	GAAGATCTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2455
OY	2341	GAAGATCTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2400
Db	2460	GAAGATCTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2519
OY	2401	TTGTGAGTCACTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2466
Db	2520	TTGTGAGTCACTGTGAGAGTGCAGTATTTTCATTGGTACCGGTGACTGATTATGGCCCTGAG	2579
OY	2461	GATTAATAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAATTAACACAGT	2520
Db	2580	GATTAATAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAATTAACACAGT	2639
OY	2521	CGGGAACAACAGCATAGAAATGGAGAAGAGTGAATGCTCACTATTTGGCAGATACA	2580
Db	2640	CGGGAACAACAGCATAGAAATGGAGAAGAGTGAATGCTCACTATTTGGCAGATACA	2699
OY	2581	TTTCAAGGTTTCAAACGCCAGTCATTTTGTCTGTTCATTAATCCAGGAATTCAGAAAG	2640
Db	2700	TTTCAAGGTTTCAAACGCCAGTCATTTTGTCTGTTCATTAATCCAGGAATTCAGAAAG	2759
OY	2641	GAATGTGCAAACTTCTGTCGCCAGTCTGGGTCCTTAAGAAACAAAGTCCAAAGTCACT	2700
Db	2760	GAATGTGCAAACTTCTGTCGCCAGTCTGGGTCCTTAAGAAACAAAGTCCAAAGTCACT	2819
OY	2701	TTTGAATGTGAACAAAGAGAAAGAAATCAAGGAAGAAATGACTTAATCAAGCCTGTA	2760
Db	2820	TTTGAATGTGAACAAAGAGAAAGAAATCAAGGAAGAAATGACTTAATCAAGCCTGTA	2879
OY	2761	CAGACAGTTAATATCTAGTCGAGGCTTTTCTGTGTGTCGAGAAAGATTAAGCCACTGAT	2820
Db	2880	CAGACAGTTAATATCTAGTCGAGGCTTTTCTGTGTGTCGAGAAAGATTAAGCCACTGAT	2939
OY	2821	AATGCCAATGTAGTATCAAAAGAGAGGCTCTAGCTTTGTCATCTCTCAGTTCAGAGGC	2880
Db	2940	AATGCCAATGTAGTATCAAAAGAGAGGCTCTAGCTTTGTCATCTCTCAGTTCAGAGGC	2999
OY	2881	AACGAACGTGACATCTATCTCCAAATTAACATGACCTTTTACAAACCCATATCGTATA	2940
Db	3000	AACGAACGTGACATCTATCTCCAAATTAACATGACCTTTTACAAACCCATATCGTATA	3059
OY	2941	CCACCATTCTTTCCCATCAAGTCATTTGTTAAACTAAATGACAAATAATCTGCTAGAG	3000
Db	3060	CCACCATTCTTTCCCATCAAGTCATTTGTTAAACTAAATGACAAATAATCTGCTAGAG	3119
OY	3001	GAAACTTTGAGGAACATTCATATGTCACCTGAAAGAGAAATGGGAAATGACAACTTCCA	3060
Db	3120	GAAACTTTGAGGAACATTCATATGTCACCTGAAAGAGAAATGGGAAATGACAACTTCCA	3179
OY	3061	AGTACAGTGAAGACATTAAGCCGTATTAACATTAGAGAAATGTTTTTAAGAGCCAGC	3120
Db	3180	AGTACAGTGAAGACATTAAGCCGTATTAACATTAGAGAAATGTTTTTAAGAGCCAGC	3239
OY	3121	TCACCAATATTTAATGAATAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTTAATGAA	3180
Db	3240	TCACCAATATTTAATGAATAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTTAATGAA	3299
OY	3181	ATAGTTCCTCACTGATGAACATTCACAGCAGACTAGTGAACACAGAGGCCAAATTTG	3240
Db	3300	ATAGTTCCTCACTGATGAACATTCACAGCAGACTAGTGAACACAGAGGCCAAATTTG	3359
OY	3241	AATGCTATGCTTAGATTTAGGGGTTTTGCAACTGAGGCTTTAAACAAGCTCTTCGGA	3300
Db	3360	AATGCTATGCTTAGATTTAGGGGTTTTGCAACTGAGGCTTTAAACAAGCTCTTCGGA	3419
OY	3301	AGTATTTTGAACATCTCTGAATTAACAAAGCAACAAATTTGAAGAAGTATGTTCAACTGTT	3360

Db	3420	AGTAATTGTAGACATCCGGAATTAATAAAGCAAGAAATATGAAGAATAGTTACAGCTGTT	3479
OY	3361	AATACAGATTTTCTCTCCATATCTGATTTTCAGATTAACATTAGACAGCCTATGGGAAGTAGT	3420
Db	3480	AATACAGATTTTCTCTCCATATCTGATTTTCAGATTAACATTAGACAGCCTATGGGAAGTAGT	3539
OY	3421	CATCAGTCTCAGGGTTGTGTCGACAGCCGATACACCTGTTAGATGATGCGTAATTAAG	3480
Db	3540	CATCAGTCTCAGGGTTGTGTCGACAGCCGATACACCTGTTAGATGATGCGTAATTAAG	3599
OY	3481	GAAGATACTATGTTTGGCAAAATGACATTTAGAGAAAGTTGCGTGTTTTATGCAAAAGC	3540
Db	3600	GAAGATACTATGTTTGGCAAAATGACATTTAGAGAAAGTTGCGTGTTTTATGCAAAAGC	3659
OY	3541	GTCCAGAAAGGAGAGCTTTAGCAGGAGTCTAGCCCTTTCACCCATACATATTTGGCTCAG	3600
Db	3660	GTCCAGAAAGGAGAGCTTTAGCAGGAGTCTAGCCCTTTCACCCATACATATTTGGCTCAG	3719
OY	3601	GGTTATCCGAAAGAGGGGCCAATAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT	3660
Db	3720	GGTTATCCGAAAGAGGGGCCAATAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT	3779
OY	3661	GAAGAGCTTCCCTGCTTCCACACACTGTTATTTGGTAAAGTAAACATATACCTTCTCAG	3720
Db	3780	GAAGAGCTTCCCTGCTTCCACACACTGTTATTTGGTAAAGTAAACATATACCTTCTCAG	3839
OY	3721	TCTACTAGGCATAGCACCGTGTGCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTTA	3780
Db	3840	TCTACTAGGCATAGCACCGTGTGCTACCGAGTGTCTGTCTAAGAACACAGAGAGAAATTTA	3899
OY	3781	TTATCATTTGAAGATAGCTTTAAATGACTGCACTAACCCAGTAAATATTGGCAAGGCATCT	3840
Db	3900	TTATCATTTGAAGATAGCTTTAAATGACTGCACTAACCCAGTAAATATTGGCAAGGCATCT	3959
OY	3841	CAGGAACATCACTCTTAGTGAGAGAAACAAATAGTCTGTACTGTTTCTTCCACAGGC	3900
Db	3960	CAGGAACATCACTCTTAGTGAGAGAAACAAATAGTCTGTACTGTTTCTTCCACAGTGC	4019
OY	3901	AGTGAATTTGGAAGACTTGACTGTCGAATTAACAACACCCAGATCCCTTCTTGATTGGTTCT	3960
Db	4020	AGTGAATTTGGAAGACTTGACTGTCGAATTAACAACACCCAGATCCCTTCTTGATTGGTTCT	4079
OY	3961	TCCAAACAAATAGAGGCATCACTCTGAAAAGCCAGGAGATGGTCTGAGTGAACAGGAATTG	4020
Db	4080	TCCAAACAAATAGAGGCATCACTCTGAAAAGCCAGGAGATGGTCTGAGTGAACAGGAATTG	4139
OY	4021	GTTTCAGATGATGAAGAAAGAGAAAGGGGCTTGGAAAGAAATATCAAGAGAGCAAGC	4080
Db	4140	GTTTCAGATGATGAAGAAAGAGAAAGGGGCTTGGAAAGAAATATCAAGAGAGCAAAAGC	4199
OY	4081	ATGGAATTCAACTTAGTGTAAGAGAGCAATCTGGGTGTGAGAGTGAACAAAGCGTCTCGAA	4140
Db	4200	ATGGAATTCAACTTAGTGTAAGAGAGCAATCTGGGTGTGAGAGTGAACAAAGCGTCTCGAA	4259
OY	4141	GACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAAACCACTCAGCAGAGGGATACCATG	4200
Db	4260	GACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAAACCACTCAGCAGAGGGATACCATG	4319
OY	4201	CAACATTAACCGATTAAGCTCCACAGAGAAATGGCTGAACCTAAGAAAGCTGTTTGAACAG	4260
Db	4320	CAACATTAACCGATTAAGCTCCACAGAGAAATGGCTGAACCTAAGAAAGCTGTTTGAACAG	4379
OY	4261	CATGGAGAGCCAGCCTTCTTAACAGCTACCCCTTCCATCAATAGTACTCTTGCCCTTGAG	4320
Db	4380	CATGGAGAGCCAGCCTTCTTAACAGCTACCCCTTCCATCAATAGTACTCTTGCCCTTGAG	4439
OY	4321	GACCTGCGAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTTCACAGAAAAGT	4380
Db	4440	GACCTGCGAAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACCTTCACAGAAAAGT	4499
OY	4381	AGTAATAATCCCTATAAAGCCAGAAATTCAGAAAGGCTTTTCTGTCGAAGTTTGAAGTGTCT	4440
Db	4500	AGTAATAATCCCTATAAAGCCAGAAATTCAGAAAGGCTTTTCTGTCGAAGTTTGAAGTGTCT	4559

QY	4441	GCAGCTACTTCTACCGAGTAAAAAATTAABAACCCAGAGTGGAAAGCTCATCCCTTCTTAA	4500
Db	4560	GCAGATAGTCTTACCGAGTAAAAAATAAACAACGAGAGTGAAGAAGTCATCCCTCTTAA	4619
QY	4501	TGCCCATCATTAAGATGATAGTGGTGCATATGCAACAGTTCTCTGGAGTCTTCAGAAATGA	4560
Db	4620	TGCCCATCATTAAGATGATAGTGGTGCATATGCAACAGTTCTCTGGAGTCTTCAGAAATGA	4679
QY	4561	AACTAACCATCTCAAGAGAGACTCATTAAGTGTGTTGATGTGAGAGCAACAGCTGGAA	4620
Db	4680	AACTAACCATCTCAAGAGAGACTCATTAAGTGTGTTGATGTGAGAGCAACAGCTGGAA	4739
QY	4621	GAGTCTGGGGCCACACGACATTTGACGGAACATCTTACTTGGCCAAGCAAGATCTAGAGGA	4680
Db	4740	GAGTCTGGGGCCACACGACATTTGACGGAACATCTTACTTGGCCAAGCAAGATCTAGAGGA	4799
QY	4681	ACCCCTTACCGTGGATCTGGAATACGCCCTCTTCGTGATGACCCGMAATCTGATCTCT	4740
Db	4800	ACCCCTTACCGTGGATCTGGAATACGCCCTCTTCGTGATGACCCGMAATCTGATCTCT	4859
QY	4741	GAAGACAGAGCCCCAGAGTACAGCTCTGTGTTGGCAACATACCATCTTCAACCTTCGATYG	4800
Db	4860	GAAGACAGAGCCCCAGAGTACAGCTCTGTGTTGGCAACATACCATCTTCAACCTTCGATYG	4919
QY	4801	AAAGTTCCTCCCAATTTGAAGTTGCAGAAATCTGCCAGAGTCCAGCTGCTCATACTACT	4860
Db	4920	AAAGTTCCTCCCAATTTGAAGTTGCAGAAATCTGCCAGAGTCCAGCTGCTCATACTACT	4979
QY	4861	GATCTGCTGGGTATTAATGCAATGGGAAGAAGTGTGAGCAGGAGGAAGCCAGATTTACA	4920
Db	4980	GATCTGCTGGGTATTAATGCAATGGGAAGAAGTGTGAGCAGGAGGAAGCCAGATTTACA	5039
QY	4921	GCTTCAACAGAAAGGCTCAACAAAAAGATGTCATGTTGGTGTCTGGCTGACCCCGAA	4980
Db	5040	GCTTCAACAGAAAGGCTCAACAAAAAGATGTCATGTTGGTGTCTGGCTGACCCCGAA	5099
QY	4981	GAATTTATGTCGTGTACAACTTTGGCCAGAAACACCAATCACTTTAATACTAATT	5040
Db	5100	GAATTTATGTCGTGTACAACTTTGGCCAGAAACACCAATCACTTTAATACTAATT	5159
QY	5041	ACTGAAGAGCTACTCATGTTGTTATGAAACAGATGCTGAGTTTGGTGTGAACGGACA	5100
Db	5160	ACTGAAGAGCTACTCATGTTGTTATGAAACAGATGCTGAGTTTGGTGTGAACGGACA	5219
QY	5101	CTGAATATTTTCTTAGAATTCGCGGAGAGAAATGGTAGTTAGCTAATTTCTGGGTCAC	5160
Db	5220	CTGAATATTTTCTTAGAATTCGCGGAGAGAAATGGTAGTTAGCTAATTTCTGGGTCAC	5279
QY	5161	CAGTCTATTTAAGAAGAAATAAGCTCATGATGACATGATTTTGAAGTCCAGAGAGATGTG	5220
Db	5280	CAGTCTATTTAAGAAGAAATAAGCTCATGATGACATGATTTTGAAGTCCAGAGAGATGTG	5339
QY	5221	GTCAATGGAAGAAACCAACAGGTCTCAAGGCGAGCAGAGAAATATCCACGAGACAGAAAGTC	5280
Db	5340	GTCAATGGAAGAAACCAACAGGTCTCAAGGCGAGCAGAGAAATATCCACGAGACAGAAAGTC	5399
QY	5281	TTCAAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCAACATGAGCCACAGATCAACTG	5340
Db	5400	TTCAAGGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCAACATGAGCCACAGATCAACTG	5459
QY	5341	GAATGGATGTACACGCTGTGTGTTGCTCTGTGTTGAAGAGACTTTCATCTTACACCTT	5400
Db	5460	GAATGGATGTACACGCTGTGTGTTGCTCTGTGTTGAAGAGACTTTCATCTTACACCTT	5519
QY	5401	GGCACAGGTGTCCACCCCAATTTGTGTTGTGACGCCAGATGCTGTGACACAGAGCAATGCG	5460
Db	5520	GGCACAGGTGTCCACCCCAATTTGTGTTGTGACGCCAGATGCTGTGACACAGAGCAATGCG	5579
QY	5461	TTTCATGCAATTTGGGACAGATGTGTGAGGACACTGTGGGACCCGAGAGTGGGTGTGAC	5520
Db	5580	TTTCATGCAATTTGGGACAGATGTGTGAGGACACTGTGGGACCCGAGAGTGGGTGTGAC	5639



QY	5521	AGTGTAGACACTCTACCACTGCGAGAGCTGGACACCTGATGATACCCAGATGCCCGAC	5589
Db	5640	AGTGTAGACACTCTACCACTGCGAGAGCTGGACACCTGATGATACCCAGATGCCCGAC	5699
OY	5581	AGCCACTAC 5589	
Db	5700	AGCCACTAC 5708	
RESULT 5	AAV46463		
ID	AAV46463	standard; cDNA; 5711 BP.	
AC	AAV46463;		
XX	18-NOV-1998	(first entry)	
DT	Human BRCA1 om12 polymorphism #6 cDNA.		
DE	BRCA1: om12; human; breast and ovarian cancer predisposing gene;		
KW	polymorphism; susceptibility; anti-oncogene; tumour suppressor;		
KW	chromosome 17q; ss.		
XX	Homo sapiens.		
OS	Key	Location/Qualifiers	
XX	FT CDS	120..5711	
XX	FT	/*tag= a	
XX	FT	/*product= "BRCA1 om12 protein"	
XX	FT	4427	
XX	FT	/*tag= b	
XX	FT	/note= "This polymorphic variation can be a T or C	
XX	XX	nucleotide"	
XX	US5750400-A.		
XX	12-MAY-1998.		
XX	12-FEB-1997;	97US-0798691.	
XX	12-FEB-1996;	96US-0598591.	
XX	12-FEB-1997;	97US-0798691.	
XX	(ONCO-) ONCORMED INC.		
XX	Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;		
XX	Schelter DB, Zeng B;		
XX	WPI: 1998-296774/26.		
XX	BRCA1 om1 gene coding sequences - useful for distinguishing between		
XX	polymorphisms and mutation(s) in the screening for disposition to		
XX	breast or ovarian cancer		
XX	Claim 2e; Page -: 54pp; English.		
XX	This sequence encodes a human BRCA1 (breast and ovarian cancer		
XX	predisposing gene) om12 gene in which a polymorphic variation occurs at		
XX	nucleotide 4427. This sequence and other polymorphic variations of this		
XX	sequence are useful for the identification of an individual who may or		
XX	may not have an increased susceptibility to breast or ovarian cancer.		
XX	The sequences used identify gene changes which are due to polymorphisms,		
XX	rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour		
XX	suppressor) which is involved in genetic inheritance of cancers,		
XX	especially breast and ovarian cancer. It is found at human chromosome		
XX	17q which is known to be linked to cancer susceptibility, especially		
XX	breast cancer. Cells containing a mutation in this gene lose the		
XX	wild-type function of BRCA1 and are more susceptible to cancers.		
XX	NOTE: This sequence does not appear in the specification but has been		
XX	created from the wild type BRCA1 om12 gene represented in AAV46449.		
XX	Sequence 5711BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;		

Query Match	100.0%	Score 5587	DB 19	Length 5711
Best Local Similarity	100.0%	Pred. No. 0		
Matches 5587	Conservative 1	Mismatches 1	Indels 0	Gaps 0
QY	1	ATGGAATTAATTCGCTTCCTTCGCGTTGMAAGATGTACAAATGTCATTAATGCTATGAGAAA	60	
Db	120	ATGGATTATTAATTCGCTTCCTTCGCGTTGMAAGATGTACAAATGTCATTAATGCTATGAGAAA	179	
QY	61	ATCTTAGAGTGTCCATCTGTCTGAGATTGATCAAGAACCTGTCTCACAAAGTGTAC	120	
Db	180	ATCTTAGAGTGTCCATCTGTCTGAGATTGATCAAGAACCTGTCTCACAAAGTGTAC	239	
QY	121	CACATATTTTGGCAATTTTGGCATGCTGGAACCTGTCCAAACGAGAAAGAGGCCCTTCACG	180	
Db	240	CACATATTTTGGCAATTTTGGCATGCTGGAACCTGTCCAAACGAGAAAGAGGCCCTTCACG	299	
QY	181	TGTCTCTTATGTAAGATGATATAACCAAAAGAGCCGTACAGAAAGTACGAGATTAGT	240	
Db	300	TGTCTCTTATGTAAGATGATATAACCAAAAGAGCCGTACAGAAAGTACGAGATTAGT	359	
QY	241	CAACTGTGTAAGAGCTATTGAAATTCATTTGTCTTTAGCTTGACACAGGTTTGAG	300	
Db	360	CAACTGTGTAAGAGCTATTGAAATTCATTTGTCTTTAGCTTGACACAGGTTTGAG	419	
QY	301	TATGCAAAAGCTATTAATTTTGCAAAAAAGAAATTAACCTCTCTGACATCTAAAGAT	360	
Db	420	TATGCAAAAGCTATTAATTTTGCAAAAAAGAAATTAACCTCTCTGACATCTAAAGAT	479	
QY	361	GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTSCCAAAAAGCTTCTACAGAT	420	
Db	480	GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTSCCAAAAAGCTTCTACAGAT	539	
QY	421	GAACCCGAAATTCCTTCTTGAGAGAAACAGTCTCAGTGTCAACTCTCTAACTTGA	480	
Db	540	GAACCCGAAATTCCTTCTTGAGAGAAACAGTCTCAGTGTCAACTCTCTAACTTGA	599	
QY	481	ACGTGAGAACTCTGAGGACCAAGAGGGGATTAACCTGAAAAGACGTCTGTACAT	540	
Db	600	ACGTGAGAACTCTGAGGACCAAGAGGGGATTAACCTGAAAAGACGTCTGTACAT	659	
QY	541	GAATTTGGATCTGATCTTCTGAGAGATACGCTTAATTAAGGCAACTTAATGTCAGTGGGA	600	
Db	660	GAATTTGGATCTGATCTTCTGAGAGATACGCTTAATTAAGGCAACTTAATGTCAGTGGGA	719	
QY	601	GATCAAGAAATTTGTACAAATCACCCCTCAGAGAACAGGGATGAATTCAGTTGGATCT	660	
Db	720	GATCAAGAAATTTGTACAAATCACCCCTCAGAGAACAGGGATGAATTCAGTTGGATCT	779	
QY	661	GCAAAAAAGCGCTGTGGAATTTTCTGAGACGATATTAACAATTAAGTCAATCATCA	720	
Db	780	GCAAAAAAGCGCTGTGGAATTTTCTGAGACGATATTAACAATTAAGTCAATCATCA	839	
QY	721	CCAGATAAATAAGATTGTAACACACACAGAGGTCACAGTACAGAGGATCCAGAAAG	780	
Db	840	CCAGATAAATAAGATTGTAACACACACAGAGGTCACAGTACAGAGGATCCAGAAAG	899	
QY	781	TATCAGGATGTTCTGTTCAAACTGCAATGTGAGGATGTGGCAACAATACATGCC	840	
Db	900	TATCAGGATGTTCTGTTCAAACTGCAATGTGAGGATGTGGCAACAATACATGCC	959	
QY	841	ACGCTATTACACATGAGAACAGAGATTTTATTAATCACTAAAGACAGAAATGAATTA	900	
Db	960	ACGCTATTACACATGAGAACAGAGATTTTATTAATCACTAAAGACAGAAATGAATTA	1019	
QY	901	AAGGCTGAATTTGTATAATAAAGCAAAACAGCTGGCTTAGCAAGAGGCCATATAACA	960	
Db	1020	AAGGCTGAATTTGTATAATAAAGCAAAACAGCTGGCTTAGCAAGAGGCCATATAACA	1079	
QY	961	TGGGCTGAAGTAAGAAACATGTAATGATAGGCGACTCCAGCACAGAAAAAAGGTA	1020	
Db	1080	TGGGCTGAAGTAAGAAACATGTAATGATAGGCGACTCCAGCACAGAAAAAAGGTA	1139	
QY	1021	GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAATTAAGCAAAACTGCCATGC	1080	



1140 GATCTGATCTGATCCCTGCTGAGAGAAAAGAAATAGCAATAGCAGAACTCCATGCG 1199  
1081 TCGAGAAATCTAGAGATGATGAGATGCTGCTGATTAACACATAATAGACATTCAG 1140  
1200 TCGAGATCTCTAGAGATGATGAGATGCTGCTGATTAACACATAATAGACATTCAG 1259  
1141 AAGCTATAGTGGTCTTCCAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1260 AAGCTATAGTGGTCTTCCAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1319  
1201 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATGATGATGATGATGATGATGAT 1260  
1320 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATGATGATGATGATGATGATGAT 1379  
1261 GAAATCTGCTGCTTCTGAGAAATAGACTTACTGAGCAGATGATGATGATGATGATGAT 1320  
1380 GAAATCTGCTGCTTCTGAGAAATAGACTTACTGAGCAGATGATGATGATGATGATGAT 1439  
1321 ATATGTAAGAGTGAAGAGTTCACCTCCAAATCAGTAGAGATATATTTGAAGACAAATA 1380  
1440 ATATGTAAGAGTGAAGAGTTCACCTCCAAATCAGTAGAGATATATTTGAAGACAAATA 1499  
1381 TTTTGGGAAAACCTATCGGAAAGAGGCAAGCCTCCCACTTAAGCCATGTAATGAAAT 1440  
1500 TTTTGGGAAAACCTATCGGAAAGAGGCAAGCCTCCCACTTAAGCCATGTAATGAAAT 1559  
1441 CTATATATAGAGATTTGTTACTGAGCAGATGATGATGATGATGATGATGATGATGAT 1500  
1560 CTATATATAGAGATTTGTTACTGAGCAGATGATGATGATGATGATGATGATGATGAT 1619  
1501 AAATTAAGGCTTAAGAGAGACCTACATCAGGCTTCATCCTGAGATTTTATCAAGAAA 1560  
1620 AAATTAAGGCTTAAGAGAGACCTACATCAGGCTTCATCCTGAGATTTTATCAAGAAA 1679  
1561 GCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATTAATCAGGAGACTAACCAAGCGAG 1620  
1680 GCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATTAATCAGGAGACTAACCAAGCGAG 1739  
1621 CAGAATGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
1740 CAGAATGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1799  
1681 TCTATTCAGATGAGAAAAATCCTAACCCAAATAGAAATCACTCGAAAAAGATCGTTTC 1740  
1800 TCTATTCAGATGAGAAAAATCCTAACCCAAATAGAAATCACTCGAAAAAGATCGTTTC 1859  
1741 AAAACGAAAGCTGACCTATTAAGCAGCAGATTAAGCAATATGCAATTCGAATTAATATC 1800  
1860 AAAACGAAAGCTGACCTATTAAGCAGCAGATTAAGCAATATGCAATTCGAATTAATATC 1919  
1801 CACAATCAAAAGACCTTAAGAAATAGGCTGAGGAGAGTCTTACAGGCAATAT 1860  
1920 CACAATCAAAAGACCTTAAGAAATAGGCTGAGGAGAGTCTTACAGGCAATAT 1979  
1861 CATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
1980 CATGCGCTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2039  
1921 ATGATATGTTGTTCTGACAGTGAAGAGATTAAGAAAAAGTAAACCAAAATGCGCAGTC 1980  
2040 ATGATATGTTGTTCTGACAGTGAAGAGATTAAGAAAAAGTAAACCAAAATGCGCAGTC 2099  
1981 AGGCACAGCAGAACTCAAACTCATGGAAGTAAAGAACCTGCAATGAGGCAAGAAAG 2040  
2100 AGGCACAGCAGAACTCAAACTCATGGAAGTAAAGAACCTGCAATGAGGCAAGAAAG 2159  
2041 AGTAACAGCAGCAATGAAGACAGATTAAGAGACATGACAGTATATCTTCCAGAGCTG 2100  
2160 AGTAACAGCAGCAATGAAGACAGATTAAGAGACATGACAGTATATCTTCCAGAGCTG 2219  
2101 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATATCCAGTAGAATTAAGAA 2160

2220 AAGTTAACAAATGACCTGCTCTTTTACTAGTGTCAATATCCAGTAGAATTAAGAA 2279  
2161 TTTGCAATCTTACCTTCCAGAGAGAAAAAGAGAAATCAGTAAGACGTTAAAGTCG 2220  
2280 TTTGCAATCTTACCTTCCAGAGAGAAAAAGAGAAATCAGTAAGACGTTAAAGTCG 2339  
2221 TCTATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTCCAAACT 2280  
2340 TCTATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTCCAAACT 2339  
2281 GAAAGATCTGTAGAGATGACGATATTTCAATTTGCTACCTGATGATTAATGAGCAGTCAG 2340  
2400 GAAAGATCTGTAGAGATGACGATATTTCAATTTGCTACCTGATGATTAATGAGCAGTCAG 2459  
2341 GAAAGTATCTGCTTACGAGAGTGAAGTGAACCTGATGAGAGAGGAAAAAGCAAAATTA 2400  
2460 GAAAGTATCTGCTTACGAGAGTGAAGTGAACCTGATGAGAGAGGAAAAAGCAAAATTA 2519  
2401 TGTGTGAGTCAAGTGTGAGCATTTGAAAAACCCCAAGGAGTAAATTCATGTTTCCAAA 2460  
2520 TGTGTGAGTCAAGTGTGAGCATTTGAAAAACCCCAAGGAGTAAATTCATGTTTCCAAA 2579  
2461 GATATATGAATATGACACAGAAAGCTTTAATGATATTCATTTGAGCATGAAATTAACACAGT 2520  
2580 GATATATGAATATGACACAGAAAGCTTTAATGATATTCATTTGAGCATGAAATTAACACAGT 2639  
2521 CGGGAACCAAGCATGAAAGTGAAGAAAGTCAATGCTGATGCTGATTTCCAAATTA 2580  
2640 CGGGAACCAAGCATGAAAGTGAAGAAAGTCAATGCTGATGCTGATTTCCAAATTA 2699  
2581 TTTCAAGTTTCAAAAGCCGACATCTTGTCTGTTTCAATCCAGAAATATGCAAGAGAG 2640  
2700 TTTCAAGTTTCAAAAGCCGACATCTTGTCTGTTTCAATCCAGAAATATGCAAGAGAG 2759  
2641 GAATGTCAACATTTCTGCCCACCTGCTGCTTTAAAGAAACAAATGCAAAAGTCACT 2700  
2760 GAATGTCAACATTTCTGCCCACCTGCTGCTTTAAAGAAACAAATGCAAAAGTCACT 2819  
2701 TTTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGATGCTAATATCAACCTGTA 2760  
2820 TTTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGATGCTAATATCAACCTGTA 2879  
2761 CAGACATTAATATCACTGACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
2880 CAGACATTAATATCACTGACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2939  
2821 AATGCCAAATGTATCAAGAGAGGCTTGAAGTTTGTCTATCATCTCAGTTCAGAGGC 2880  
2940 AATGCCAAATGTATCAAGAGAGGCTTGAAGTTTGTCTATCATCTCAGTTCAGAGGC 2999  
2881 AAGCAACTGGACATTAATCTCCAAATTAACATGAGCTTTTACAAACCCATATCGTATA 2940  
3000 AAGCAACTGGACATTAATCTCCAAATTAACATGAGCTTTTACAAACCCATATCGTATA 3059  
2941 CCAACACTTTTCCCATCAAGTCAATTTGTTAAACCTAAATGAAGAAAAATCTCTAGAG 3000  
3060 CCAACACTTTTCCCATCAAGTCAATTTGTTAAACCTAAATGAAGAAAAATCTCTAGAG 3119  
3001 GAAACCTTTGAGAGCAATTTCAATGTCACTGAAAGAGAAATGGGAAATGACACATTTCA 3060  
3120 GAAACCTTTGAGAGCAATTTCAATGTCACTGAAAGAGAAATGGGAAATGACACATTTCA 3179  
3061 AGTACAGTGAAGCAATTTAGCCGTAATTAATAGAGAAATGTTTAAAGAGCCAGC 3120  
3180 AGTACAGTGAAGCAATTTAGCCGTAATTAATAGAGAAATGTTTAAAGAGCCAGC 3239  
3121 TCAAGCAATATTTATGAGTATGAGTTCAGTACTAATGAATGAGGCTCAGTATTAATGA 3180  
3240 TCAAGCAATATTTATGAGTATGAGTTCAGTACTAATGAATGAGGCTCAGTATTAATGA 3239  
3181 ATAGGTTCCAGTATGAGAAACATTTCAAGCAAGAACTAGTGAAGAACAGAGGCCCCAAATTTG 3240  
3300 ATAGGTTCCAGTATGAGAAACATTTCAAGCAAGAACTAGTGAAGAACAGAGGCCCCAAATTTG 3359

3241 AATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTATATAACAAGCTCTCCGGA 3300  
|||||  
3360 AATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTATATAACAAGCTCTCCGGA 3419  
|||||  
3301 AGTAATTTGAAGCATCTGAAATTAACCAAGAAATATGAAGAAGTAGTTCAGACTGT 3360  
|||||  
3420 AGTAATTTGAAGCATCTGAAATTAACCAAGAAATATGAAGAAGTAGTTCAGACTGT 3479  
|||||  
3361 AATACGATTTCTCTCCATATCTGATTTTCAGATTAACCAAGCTATGGGAAGTAGT 3420  
|||||  
3480 AATACGATTTCTCTCCATATCTGATTTTCAGATTAACCAAGCTATGGGAAGTAGT 3539  
|||||  
3421 CATGCTCTCAGGTTGTTCTGAGACACGCTGATGACCTGTAGATGATGATGAATTAAG 3480  
|||||  
3540 CATGCTCTCAGGTTGTTCTGAGACACGCTGATGACCTGTAGATGATGATGAATTAAG 3539  
|||||  
3481 GAAGATAGTAGTTTCTCTGAAATGACATTAAGGAAAGTTCTGCTGTTTTCAGCAAAAGC 3540  
|||||  
3600 GAAGATAGTAGTTTCTCTGAAATGACATTAAGGAAAGTTCTGCTGTTTTCAGCAAAAGC 3659  
|||||  
3541 GTCCAGAAAGAGAGGCTTGAAGAGCTGAGGCTTTCACCCATACACATTTGGGCTCAG 3600  
|||||  
3660 GTCCAGAAAGAGAGGCTTGAAGAGCTGAGGCTTTCACCCATACACATTTGGGCTCAG 3719  
|||||  
3601 GGTACCGAAGAGGGGCGCAAGAAATTAAGAGTCCCTCAGAAAGAACTTATCTAGTAGAGAT 3660  
|||||  
3720 GGTACCGAAGAGGGGCGCAAGAAATTAAGAGTCCCTCAGAAAGAACTTATCTAGTAGAGAT 3779  
|||||  
3661 GAAGAGCTTCCCTGCTTCCAAACCTGTTATTTGGTAAATAAACAATATACCTTCTCAG 3720  
|||||  
3780 GAAGAGCTTCCCTGCTTCCAAACCTGTTATTTGGTAAATAAACAATATACCTTCTCAG 3839  
|||||  
3721 TCTACTAGGCTATAGACCGCTGCTACCGAGTGTCTGTAGAAACAGAGGGAATTTA 3780  
|||||  
3840 TCTACTAGGCTATAGACCGCTGCTACCGAGTGTCTGTAGAAACAGAGGGAATTTA 3899  
|||||  
3781 TTATCATTTGAAGAAATAGCTTAATTAATGACTGCACTAACAGTAATTTGGCAAGGCAATCT 3840  
|||||  
3900 TTATCATTTGAAGAAATAGCTTAATTAATGACTGCACTAACAGTAATTTGGCAAGGCAATCT 3959  
|||||  
3841 CAGGAACATCACCCTTATGAGAGAACAAATGTTGCTGTAAGCTTGTCTTCTACAGTGC 3900  
|||||  
3960 CAGGAACATCACCCTTATGAGAGAACAAATGTTGCTGTAAGCTTGTCTTCTACAGTGC 4019  
|||||  
3901 AGTGAATTTGAAGAACTGACTGCAATTAACCAACCCAGAGATCTTCTGATTGGTCT 3960  
|||||  
4020 AGTGAATTTGAAGAACTGACTGCAATTAACCAACCCAGAGATCTTCTGATTGGTCT 4079  
|||||  
3961 TCCAAACAATGAGGCTAGTCTGAAAGCCAGAGGATGCTGTAAGTGAACAAGAAATG 4020  
|||||  
4080 TCCAAACAATGAGGCTAGTCTGAAAGCCAGAGGATGCTGTAAGTGAACAAGAAATG 4139  
|||||  
4021 GTTTTCAGATGATGAAGAAAGAGAACCGGCTTGAAGAAATTAATCAAGAAGCAAAAGC 4080  
|||||  
4140 GTTTTCAGATGATGAAGAAAGAGAACCGGCTTGAAGAAATTAATCAAGAAGCAAAAGC 4199  
|||||  
4081 ATGGAATTTCAACTTAGTAGAGCAGCATCTGGGTGAGAGTGAACAACCGCTCTGAA 4140  
|||||  
4200 ATGGAATTTCAACTTAGTAGAGCAGCATCTGGGTGAGAGTGAACAACCGCTCTGAA 4259  
|||||  
4141 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAAACCACTCAGAGAGAGGATACCATG 4200  
|||||  
4260 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAAACCACTCAGAGAGAGGATACCATG 4319  
|||||  
4201 CAACATTAACCTGATTAAGCTCCAGAGAAATGCTGAATAGTAGAGCTGTGTTAAGAACAG 4260  
|||||  
4320 CAACATTAACCTGATTAAGCTCCAGAGAAATGCTGAATAGTAGAGCTGTGTTAAGAACAG 4379  
|||||  
4261 CATGGAGGAGGCTCTTAACAGCTACCTTCATATAGTAGTACGCTTCTGCCCTTGAG 4320  
|||||  
4380 CATGGAGGAGGCTCTTAACAGCTACCTTCATATAGTAGTACGCTTCTGCCCTTGAG 4439  
|||||

4321 GACTCGCAAAATCCAGAACACATCAGAGAAAGACAGTATTAATCTTCAAGAAAGT 4380  
|||||  
4440 GACTCGCAAAATCCAGAACACATCAGAGAAAGACAGTATTAATCTTCAAGAAAGT 4499  
|||||  
4381 AGTGAATTTCAACTTAGTAGAGCAGATCCAGAGGCTTGTGCTGACAAAGTTGAGGTGCT 4440  
|||||  
4500 AGTGAATTTCAACTTAGTAGAGCAGATCCAGAGGCTTGTGCTGACAAAGTTGAGGTGCT 4559  
|||||  
4441 GCAGATAGTCTCCAGTAAATAAAGAACCGAGAGTGAAGAAAGGATCAATCCCTCTAA 4500  
|||||  
4560 GCAGATAGTCTCCAGTAAATAAAGAACCGAGAGTGAAGAAAGGATCAATCCCTCTAA 4619  
|||||  
4501 TGCCCATCATTTAGATATAGTGTGTACATGACAGTGTGCTGAGAGTCTCAGAAATGA 4560  
|||||  
4620 TGCCCATCATTTAGATATAGTGTGTACATGACAGTGTGCTGAGAGTCTCAGAAATGA 4679  
|||||  
4561 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAGAGAGCAACAGCTGGA 4620  
|||||  
4680 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAGAGAGCAACAGCTGGA 4739  
|||||  
4621 GAGTCTGGGCGCACAGATTTGACGGAACATCTTCTGCAAGGCAAGATCTAGAGGA 4680  
|||||  
4740 GAGTCTGGGCGCACAGATTTGACGGAACATCTTCTGCAAGGCAAGATCTAGAGGA 4799  
|||||  
4681 ACCCTTACCTGGAATCTGGAATCAGCCTTCTCTGATGACCTGGAATCTGATCTTCT 4740  
|||||  
4800 ACCCTTACCTGGAATCTGGAATCAGCCTTCTCTGATGACCTGGAATCTGATCTTCT 4859  
|||||  
4741 GAAGACAGAGCCCCAGAGTACGCTGTGTTGGCAACATACCATTTCAACCTCTGCATTG 4800  
|||||  
4860 GAAGACAGAGCCCCAGAGTACGCTGTGTTGGCAACATACCATTTCAACCTCTGCATTG 4919  
|||||  
4801 AAAGTTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCCAGAGCTGCTCATCTACT 4860  
|||||  
4920 AAAGTTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCCAGAGCTGCTCATCTACT 4979  
|||||  
4861 GATACGCTGGGTATATGCAATGAAGAAGTGTGAGGAGGAGAGGACCAAAATTTGACA 4920  
|||||  
4980 GATACGCTGGGTATATGCAATGAAGAAGTGTGAGGAGGAGAGGACCAAAATTTGACA 5039  
|||||  
4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGTGTGTGCTGCTGACCCAGAA 4980  
|||||  
5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGTGTGTGCTGCTGACCCAGAA 5099  
|||||  
4981 GAATTTATGCTGTGTGACAAAGTTTCCAGAAACACACATCACTTAACTAATTAAT 5040  
|||||  
5100 GAATTTATGCTGTGTGACAAAGTTTCCAGAAACACACATCACTTAACTAATTAAT 5139  
|||||  
5041 ACTGAAGAGACTACTGATGTTGTTATGAAGACAGATGCTGACTTGTGTGAACGAGCA 5100  
|||||  
5160 ACTGAAGAGACTACTGATGTTGTTATGAAGACAGATGCTGACTTGTGTGAACGAGCA 5219  
|||||  
5101 CTGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGACC 5160  
|||||  
5220 CTGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGACC 5279  
|||||  
5161 CACTTATTTAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGCAATGTG 5220  
|||||  
5280 CACTTATTTAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGCAATGTG 5339  
|||||  
5221 GTCAATGAAAGAAACCCAGAGTCCAAAGCAGAGCAAGAAATCCAGAGCAAGAAATG 5280  
|||||  
5340 GTCAATGAAAGAAACCCAGAGTCCAAAGCAGAGCAAGAAATCCAGAGCAAGAAATG 5399  
|||||  
5281 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAACATGCCCCAGATCACTG 5340  
|||||  
5400 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAACATGCCCCAGATCACTG 5459  
|||||  
5341 GAATGATGTTGACAGTGTGTGCTGCTGTTGGAAGGAGGACCTTCAATCACTCCCTT 5400  
|||||  
5460 GAATGATGTTGACAGTGTGTGCTGCTGTTGGAAGGAGGACCTTCAATCACTCCCTT 5519  
|||||  
5401 GGCACAGGTGCTCCACCAATTTGTTGTGACAGCAGATGCTTGACAGAGACAAATGCG 5460  
|||||

Db		5520	GGCAGAGTGTCCACCAATTTGTTGTGTGCAGGCACATGCTTGACAGAGACAATGCC	5579
Oy		5461	TTCATCATTAATTGGGCGAGATGTGTGAGCGACACTGTGTGACCAGAGTGGGTGTGGAC	5520
Db		5580	TTCCATCATTAATTGGGCGAGATGTGTGAGCGACACTGTGTGACCCGAGAGTGGGTGTGGAC	5639
Oy		5521	AGTTAGCACACTTACCCAGTGTGCAGAGACTTGACACCTACCTGAATACCCAGATCCCCAC	5580
Db		5640	AGTTAGCACACTTACCCAGTGTGCAGAGACTTGACACCTACCTGAATACCCAGATCCCCAC	5699
Oy		5581	AGCCACTAC 5589	
Db		5700	AGCCACTAC 5708	
RESULT 6				
ID	AAV46464	standard; cDNA; 5711 BP.		
XX	AAV46464;			
AC	AAV46464;			
XX				
DT	18-NOV-1998	(first entry)		
XX				
DE	Human BRCA1 omi2 polymorphism #7 CDNA.			
KW	BRCA1; omi1; human; breast and ovarian cancer predisposing gene;			
KM	polymorphism; susceptibility; anti-oncogene; tumour suppressor;			
KW	chromosome 17q; ss.			
XX				
OS	Homo sapiens.			
XX				
FT	Key	Location/Qualifiers		
FT	CDS	120..5711		
FT		/tag= a		
FT		/product= "BRCA1 omi2 protein"		
FT	variation	4956		
FT		/tag= b		
FT		/note= "this polymorphic variation can be an A or G nucleotide"		
PX	US5750400-A.			
PN				
XX	12-MAY-1998.			
XX				
PF	12-FEB-1997;	97US-0798691.		
XX				
PR	12-FEB-1996;	96US-0598591.		
PR	12-FEB-1997;	97US-0798691.		
XX				
PA	(ONCO-) ONCORMED INC.			
PI	Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;			
PI	Scheider DB, Zeng B;			
DR	WPI: 1998-296774/26.			
XX				
PT	BRCA1 omi gene coding sequences - useful for distinguishing between			
PT	polymorphisms and mutation(s) in the screening for disposition to			
PT	breast or ovarian cancer			
XX				
PS	Claim 2e; Page -: 54pp; English.			
CC	This sequence encodes a human BRCA1 (breast and ovarian cancer			
CC	predisposing gene) omi2 gene in which a polymorphic variation occurs at			
CC	nucleotide 4956. This sequence and other polymorphic variations of this			
CC	sequence are useful for the identification of an individual who may or			
CC	may not have an increased susceptibility to breast or ovarian cancer.			
CC	The sequences used identify gene changes which are due to polymorphisms,			
CC	rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour			
CC	suppressor) which is involved in genetic inheritance of cancers,			
CC	especially breast and ovarian cancer. It is found at human chromosome			
CC	17q which is known to be linked to cancer susceptibility, especially			

Query Match	100.0%	Score 5587	DB 19	Length 5711
Best Local Similarity	100.0%	Pred. No. 0		
Matches 5587	Conservative 1	Mismatches 1	Indels 0	Gaps 0
OY	1	ATGATTTATTCGCTTCGGGTTGAAGAAAGTACAAATATGCTATTAATGCTATGACGAAA	60	
Db	120	ATGATTTATTCGCTTCGGGTTGAAGAAAGTACAAATATGCTATTAATGCTATGACGAAA	179	
OY	61	ATCTTAGAGTGTCCCATCTGTCGTGAGTTGGATCAAGAAACCTGTCTCCACAAGAGTGAGC	120	
Db	180	ATCTTAGAGTGTCCCATCTGTCGTGAGTTGGATCAAGAAACCTGTCTCCACAAGAGTGAGC	239	
OY	121	CACATATTTGCAAAATTTTGCATGCTGAAACTTCTCAACACAGACAGAAAGGCGCTTCACAG	180	
Db	240	CACATATTTTCCAAATTTTGGATGCTGAAACTTCTCAACACAGAAAGGCGCTTCACAG	299	
OY	181	TGTCCTTATGTAGAAATGATATATACCAAAAGSAGCCATCAAGAAAGTACGAGATTTCAT	240	
Db	300	TGTCCTTATGTAGAAATGATATATACCAAAAGSAGCCATCAAGAAAGTACGAGATTTCAT	359	
OY	241	CAACTGTGTTGAAGAGCTAFTTGAATAATCAATTTGTGCTTTTCAGCTTGACACAGATTGGAG	300	
Db	360	CAACTGTGTTGAAGAGCTAFTTGAATAATCAATTTGTGCTTTTCAGCTTGACACAGATTGGAG	419	
OY	301	TATGCAAAACAGCTATATATTTTGGCAAAAAGGAAATAACTCTCTGCAACATCTAAAGAT	360	
Db	420	TATGCAAAACAGCTATATATTTTGGCAAAAAGGAAATAACTCTCTGCAACATCTAAAGAT	479	
OY	361	GAGCTTCTATCATCCCAAGTATATGGCTATACAGAAACCGTGGCAAAAGACTTCTACAGAT	420	
Db	480	GAGCTTCTATCATCCCAAGTATATGGCTATACAGAAACCGTGGCAAAAGACTTCTACAGAT	539	
OY	421	GAACCCGAAATTCCTCCCTTGACAGAAACAGCTCAGTGTGCCAAGCTCTAACCTTGA	480	
Db	540	GAACCCGAAATTCCTCCCTTGACAGAAACAGCTCAGTGTGCCAAGCTCTAACCTTGA	599	
OY	481	ACTGTGAGACCTGTGAGGACAAAGCAGCGATACAAACCTCAAAAGACGTTGCTACATT	540	
Db	600	ACTGTGAGACCTGTGAGGACAAAGCAGCGATACAAACCTCAAAAGACGTTGCTACATT	659	
OY	541	GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATTAAGGCACTTATTTGCAGTGGGA	600	
Db	660	GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATTAAGGCACTTATTTGCAGTGGGA	719	
OY	601	GATCAAGAATTTGTACAAATCACCCCTCAAGGAACCAAGGATGAATCAGTTGGATTCT	660	
Db	720	GATCAAGAATTTGTACAAATCACCCCTCAAGGAACCAAGGATGAATCAGTTGGATTCT	779	
OY	661	GCAAAAAAGCGTGTGTGAATTTTCTGAGACGAGTGTACAAATATCTGAACATCATCAA	720	
Db	780	GCAAAAAAGCGTGTGTGAATTTTCTGAGACGAGTGTACAAATATCTGAACATCATCAA	839	
OY	721	CCCAGTAAATATGATTTTGAACACCACTGAGAAAGCGTGCACGTGAGAGCATCCAGAAAG	780	
Db	840	CCCAGTAAATATGATTTTGAACACCACTGAGAAAGCGTGCACGTGAGAGCATCCAGAAAG	899	
OY	781	TATCAGGAGTATCTGTTTCAAACTGTGATGGGAGCATGGGACAAATCTCATGCG	840	
Db	900	TATCAGGAGTATCTGTTTCAAACTGTGATGGGAGCATGGGACAAATCTCATGCG	959	
OY	841	AGCTCATTTACAGATGAGACAGCAGTTTATTACTACTAAAGACAGATGATGTAGAA	900	
Db	960	AGCTCATTTACAGATGAGACAGCAGTTTATTACTACTAAAGACAGATGATGTAGAA	1019	
OY	901	AAGGCTGAATTTCTGTAATTAAGCAACAGCGCTGCTTATGACAGGACCAACATATACGA	960	

Db 1020 AAGGCTGAATTCTGTATATAAACCAACACGCTGGCTTAGCAAGAGCCCAACATTAACAGA 1079  
QY 961 TGGGCTGSAAGTAAGGAAACATGTAATAGTAGGGGGCTCCCGACACAGAAAAAAGSTA 1020  
Db 1080 TGGGCTGGAAGTAAGGAAACATGTAATAGTAGGGGGCTCCCGACACAGAAAAAAGSTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGGAATTAAGCAAACTGCCATGC 1080  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGGAATTAAGCAAACTGCCATGC 1139  
QY 1081 TCAGAGAAATCTTAGAGATCTGAAGATGTTCTTGGAATAACTAAATAGCAGATTGAG 1140  
Db 1200 TCAGAGAAATCTTAGAGATCTGAAGATGTTCTTGGAATAACTAAATAGCAGATTGAG 1259  
QY 1141 AAGGTAATGAGAGGTTTTCCAGAGAGATGAAGTGAAGTTCTGTGATCAGATCAGAT 1200  
Db 1260 AAGGTAATGAGAGGTTTTCCAGAGAGATGAAGTGAAGTTCTGTGATCAGATCAGAT 1319  
QY 1201 GGGGAGCTGGAATCAATGCCAAAGTAGCTGATGTAATGGACGTTCTTAAATGAGGTAGAT 1260  
Db 1320 GGGGAGCTGGAATCAATGCCAAAGTAGCTGATGTAATGGACGTTCTTAAATGAGGTAGAT 1379  
QY 1261 GAATATTCCTGCTTCTCAGAGAAATAGACTTACTGCGCCAGTATCTCTATGAGGCTTTA 1320  
Db 1380 GAATATTCCTGCTTCTCAGAGAAATAGACTTACTGCGCCAGTATCTCTATGAGGCTTTA 1439  
QY 1321 ATATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1380  
Db 1440 ATATGTAAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1499  
QY 1381 TTTGGAAAACTTATCGGAAGAGCAAGCCCTCCCAACTTAAGCCATGTAACTGAAAT 1440  
Db 1500 TTTGGAAAACTTATCGGAAGAGCAAGCCCTCCCAACTTAAGCCATGTAACTGAAAT 1559  
QY 1441 CTATATATGAGACATTTGTTACTGAGCCACAGATATATAAGAGCCTCCCTCACAAT 1500  
Db 1560 CTATATATGAGACATTTGTTACTGAGCCACAGATATATAAGAGCCTCCCTCACAAT 1619  
QY 1501 AATATTAAGCGTAAAGAGAGACTTACATGAGCCCTTCTGATGAGATTTTATCAAGAA 1560  
Db 1620 AATATTAAGCGTAAAGAGAGACTTACATGAGCCCTTCTGATGAGATTTTATCAAGAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAAGACTCCTGAATGATAAATCAGGAGTAACCAACGAG 1620  
Db 1680 GCAGATTTGGCAGTTCAAAAAGACTCCTGAATGATAAATCAGGAGTAACCAACGAG 1739  
QY 1621 CAGAAATGTCAGATGATGATATTACTAATAGTGTGATGAGAAATAAAGAGTGAT 1680  
Db 1740 CAGAAATGTCAGATGATGATATTACTAATAGTGTGATGAGAAATAAAGAGTGAT 1799  
QY 1681 TCTATTCAGATGAGAAAAATCCTTAACCAATAGAAATCAGTCGAAAAAAGATGCTGTTTC 1740  
Db 1800 TCTATTCAGATGAGAAAAATCCTTAACCAATAGAAATCAGTCGAAAAAAGATGCTGTTTC 1859  
QY 1741 AAAAGCAAGCTGAGACCTTATAGCAGAGATATAGCAATATGGAATCGAATTTAATATC 1800  
Db 1860 AAAAGCAAGCTGAGACCTTATAGCAGAGATATAGCAATATGGAATCGAATTTAATATC 1919  
QY 1801 CACAATTTCAAAAGCAGCTAAAGAAATAGGCTGAGAGGAAAGCTTCTACAGGCAATAT 1860  
Db 1920 CACAATTTCAAAAGCAGCTAAAGAAATAGGCTGAGAGGAAAGCTTCTACAGGCAATAT 1979  
QY 1861 CATGCGCTTGAAGTACTAGCTAGTAAGAAATCTAAGCCACCTAATGTACTGAATTCGA 1920  
Db 1980 CATGCGCTTGAAGTACTAGCTAGTAAGAAATCTAAGCCACCTAATGTACTGAATTCGA 2039  
QY 1921 ATTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGATACAAACCAATGCCAGTC 1980  
Db 2040 ATTGATAGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGATACAAACCAATGCCAGTC 2039  
QY 1981 AGGCACAGCAAGAACTTACAACTCATGTGAAGGTAAGAGACCTGCAACTGAGCCAGAG 2040  
Db 2100 AGGCACAGCAAGAACTTACAACTCATGTGAAGGTAAGAGACCTGCAACTGAGCCAGAG 2159  
QY 2041 AGTACAGCCCAATGAGACAGACAGTAAAGACATGACAGTATCTTTCCAGAGCTG 2100  
Db 2160 AGTACAGCCCAATGAGACAGACAGTAAAGACATGACAGGATCTTTCCAGAGCTG 2219  
QY 2101 AAGTTAACAATGCACCTGGTCTTTTACTAAGTGTTCAAATACAGTAACCTTAAGAA 2160  
Db 2220 AAGTTAACAATGCACCTGGTCTTTTACTAAGTGTTCAAATACAGTAACCTTAAGAA 2279  
QY 2161 TTTGTCAATCCTAGCCTTCCAGAGAGAAAAAAGAGACAACTAGAAACAGTTAAAGTG 2220  
Db 2280 TTTGTCAATCCTAGCCTTCCAGAGAGAAAAAAGAGACAACTAGAAACAGTTAAAGTG 2339  
QY 2221 TCTAATATGCTGAGAGACCCCAAGATTCATGTTAAGTGGAGAAAGGTTTTGCAAACT 2280  
Db 2340 TCTAATATGCTGAGAGACCCCAAGATTCATGTTAAGTGGAGAAAGGTTTTGCAAACT 2399  
QY 2281 GAAAGATCTGTAGAGATGACAGTATTTCAATTTGTAAGTACGTATTAATGAGCAGTCAG 2340  
Db 2400 GAAAGATCTGTAGAGATGACAGTATTTCAATTTGTAAGTACGTATTAATGAGCAGTCAG 2459  
QY 2341 GAAAGTATCTGTTACTGGAAGTTAGCAGCTTAGGGAAGGCAAAAAACAACTAATAA 2400  
Db 2460 GAAAGTATCTGTTACTGGAAGTTAGCAGCTTAGGGAAGGCAAAAAACAACTAATAA 2519  
QY 2401 TGTGTGAGTCACTGTCAGACATTTGAAAAACCCCAAGGACTAATTCATGTTGTTCCAA 2460  
Db 2520 TGTGTGAGTCACTGTCAGACATTTGAAAAACCCCAAGGACTAATTCATGTTGTTCCAA 2579  
QY 2461 GATATATGAATATGACACAGAGGCTTTAAGTATCCATTTGGGACATGAATTAACACAGT 2520  
Db 2580 GATATATGAATATGACACAGAGGCTTTAAGTATCCATTTGGGACATGAATTAACACAGT 2639  
QY 2521 CGGGAACAAGCATATGAGAAATGGAAGAAAGAACTGATGCTCAGATTTGAGATACA 2580  
Db 2640 CGGGAACAAGCATATGAGAAATGGAAGAAAGAACTGATGCTCAGATTTGAGATACA 2699  
QY 2581 TTTCAAGGTTTCAAAAGCGCCAGTCTATTTGCTTTTCAAAATCCAGGAATCCAGAGAG 2640  
Db 2700 TTTCAAGGTTTCAAAAGCGCCAGTCTATTTGCTTTTCAAAATCCAGGAATCCAGAGAG 2759  
QY 2641 GAATGTGCAACATTTCTGCCCCACTGTGGCTTTAAGCAAAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGCAACATTTCTGCCCCACTGTGGCTTTTAAAGAAAGTCCAAAGTCACT 2819  
QY 2701 TTTGAATGTGACAAAAAGAAAGAAATCAAGGAAAGAAATGAGCTAATTTCAAGCTGTA 2760  
Db 2820 TTTGAATGTGACAAAAAGAAAGAAATCAAGGAAAGAAATGAGCTAATTTCAAGCTGTA 2879  
QY 2761 CAGACAGTTAATATCACTGAGGCTTTCTGTGTTGTTGTCAGAAAGATTAAGCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCACTGAGGCTTTCTGTGTTGTTGTCAGAAAGATTAAGCAGTTGAT 2939  
QY 2821 AATGCCAATGTATGATCAAGAGAGCTCTAGGTTTGTCTATCATCTCAGTTCCAGAGC 2880  
Db 2940 AATGCCAATGTATGATCAAGAGAGCTCTAGGTTTGTCTATCATCTCAGTTCCAGAGC 2999  
QY 2881 AAGCAAGCTGAGACTCTTCAATTAACATTAAGCAATGAGCTTTCAAAACCCATATGTA 2940  
Db 3000 AAGCAAGCTGAGACTCTTCAATTAACATTAAGCAATGAGCTTTCAAAACCCATATGTA 3059  
QY 2941 CCACACACTTTTCCATCAAGTCAATTTGTTAAACATTAATGTAAAGAAAAATCTGCTAGAG 3000  
Db 3060 CCACACACTTTTCCATCAAGTCAATTTGTTAAACATTAATGTAAAGAAAAATCTGCTAGAG 3119  
QY 3001 GAAACTTTGAGGACATTCATATGTCACCTGAAAGAGAAATGGAATGAGAACATTTCCA 3060  
Db 3120 GAAACTTTGAGGACATTCATATGTCACCTGAAAGAGAAATGGAATGAGAACATTTCCA 3179  
QY 3061 AGTACAGTGAAGCAATTTAGCCGTAAATTAACATTTAGAGAAAAATGTTTTTAAGAACCCAGC 3120  
Db 3180 AGTACAGTGAAGCAATTTAGCCGTAAATTAACATTTAGAGAAAAATGTTTTTAAGAACCCAGC 3239

QY 3121 TCAGCAATTTATTAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3180  
DB 3240 TCAGCAATTTATTAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3299  
QY 3181 ATAGGTTCCAGTATGAAGCAATTCAGAGAACTGATAGAAACGAGAGGCCAAATG 3240  
DB 3300 ATAGGTTCCAGTATGAAGCAATTCAGAGAACTGATAGAAACGAGAGGCCAAATG 3359  
QY 3241 AATGCTATGCTTATGATTAAGGGGTTTTCAGACCTGAGGCTATTAACAAAGTCTCTGGA 3300  
DB 3360 AATGCTATGCTTATGATTAAGGGGTTTTCAGACCTGAGGCTATTAACAAAGTCTCTGGA 3419  
QY 3301 AGTAATTTGATGATTCCTGAATTAAGCAAGATATGAAGAGTATGACAGTCT 3360  
DB 3420 AGTAATTTGATGATTCCTGAATTAAGCAAGATATGAAGAGTATGACAGTCT 3479  
QY 3361 AATACGATTTCTCTCCATATCTGATTTAGATTAATTAGAACGCTTATGGAATAGT 3420  
DB 3480 AATACGATTTCTCTCCATATCTGATTTAGATTAATTAGAACGCTTATGGAATAGT 3539  
QY 3421 CATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTTAGATGATGATGAATTAAG 3480  
DB 3540 CATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTTAGATGATGATGAATTAAG 3599  
QY 3481 GAAGATAGTATGTTTCTGAAAAATGACATTAAAGAAATGCTGCTGTTTTCAGCAAGC 3540  
DB 3600 GAAGATAGTATGTTTCTGAAAAATGACATTAAAGAAATGCTGCTGTTTTCAGCAAGC 3659  
QY 3541 GTCCAAAGAGAGGCTTATGACAGAGTCTTACGCTTACCCATACATTTGGCTCAG 3600  
DB 3660 GTCCAAAGAGAGGCTTATGACAGAGTCTTACGCTTACCCATACATTTGGCTCAG 3719  
QY 3601 GGTTCAGGAAGAGGGGCCAAGAAATAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT 3660  
DB 3720 GGTTCAGGAAGAGGGGCCAAGAAATAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT 3779  
QY 3661 GAAGAGCTTCCCTGCTTCCAAACATTTGTAATTTGTAAGTAACAATATACCTTCTCAG 3720  
DB 3780 GAAGAGCTTCCCTGCTTCCAAACATTTGTAATTTGTAAGTAACAATATACCTTCTCAG 3839  
QY 3721 TCTACTAGGCTATGACGCTGCTTACGAGTCTCTTAAGAACACAGAGGAATTTA 3780  
DB 3840 TCTACTAGGCTATGACGCTGCTTACGAGTCTCTCTTAAGAACACAGAGGAATTTA 3899  
QY 3781 TTATCTTGAAGAAATGCTTAAATGACTGACAGTAACAGGATTAATTTGGCAAGGATCT 3840  
DB 3900 TTATCTTGAAGAAATGCTTAAATGACTGACAGTAACAGGATTAATTTGGCAAGGATCT 3959  
QY 3841 CAGGAACATCACCTTATGAGAGAAACAAATGTTCTGCTAGCTTGTCTTCTCACAGTGC 3900  
DB 3960 CAGGAACATCACCTTATGAGAGAAACAAATGTTCTGCTAGCTTGTCTTCTCACAGTGC 4019  
QY 3901 AATGATTTGGAAGACTTGACTGCTCAATTAACAACCCAGGATCTTCTTGTGATGTTCT 3960  
DB 4020 AATGATTTGGAAGACTTGACTGCTCAATTAACAACCCAGGATCTTCTTGTGATGTTCT 4079  
QY 3961 TCCAAACAATGAGGATGATCTGTAAGCCAGAGGAGTGTCTGATGACAAAGAAATG 4020  
DB 4080 TCCAAACAATGAGGATGATCTGTAAGCCAGAGGAGTGTGTCTGATGACAAAGAAATG 4139  
QY 4021 GTTTCAGATGATGAAGAAAGAGAACGGGCTTGAAGAAATATCAAGAGAGCAAGC 4080  
DB 4140 GTTTCAGATGATGAAGAAAGAGAACGGGCTTGAAGAAATATCAAGAGAGCAAGC 4199  
QY 4081 ATGGAATCAACTTGAAGTGAAGCAGCATGTTGGTGTGAAGTGAACAGGCTCTGAA 4140  
DB 4200 ATGGAATCAACTTGAAGTGAAGCAGCATGTTGGTGTGAAGTGAACAGGCTCTGAA 4259  
QY 4141 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAACACTCAGAGAGGATATCAGATG 4200  
DB 4260 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAACACTCAGAGAGGATATCAGATG 4319  
QY 4201 CAACATTAACCTGATTAAGCTCCAGAGAAATGCTGAACTAGAAAGCTGTGTAGAAAG 4260

DB 4320 CAACATTAACCTGATTAAGCTCCAGAGAAATGCTGAACTAGAAAGCTGTGTAGAAAGCAG 4379  
QY 4261 CATGGAGGAGGCTTCTTAACACTACCTTCCATCTAATGATGACCTTGTGCTTGAAG 4320  
DB 4380 CATGGAGGAGGCTTCTTAACACTACCTTCCATCTAATGATGACCTTGTGCTTGAAG 4439  
QY 4321 GACCTGGAATTCAGAACCAAGCACATCCAGAAAAAGCACTATTAATCTTCAGAAAGT 4380  
DB 4440 GACCTGGAATTCAGAACCAAGCACATCCAGAAAAAGCACTATTAATCTTCAGAAAGT 4499  
QY 4381 AGTGAATTAACCTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTGAGGTCT 4440  
DB 4500 AGTGAATTAACCTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAAGTTGAGGTCT 4559  
QY 4441 GCAGATAGTTCTACAGTAAATTAAGAACCAAGAGTGAAGAGTCAATCCCTTCTAA 4500  
DB 4560 GCAGATAGTTCTACAGTAAATTAAGAACCAAGAGTGAAGAGTCAATCCCTTCTAA 4619  
QY 4501 TGCCCATCATTAATGATAGTGTGATACATGACAGTTGCTGAGGATCTTCAGAAATGA 4560  
DB 4620 TGCCCATCATTAATGATAGTGTGATACATGACAGTTGCTGAGGATCTTCAGAAATGA 4679  
QY 4561 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAGAGCAACAGCTGAA 4620  
DB 4680 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAGAGCAACAGCTGAA 4739  
QY 4621 GAGTGGGGCCACAGATTTGAGGAAACATCTACTTGGCAAGGCAAGTCTAGAGGA 4680  
DB 4740 GAGTGGGGCCACAGATTTGAGGAAACATCTACTTGGCAAGGCAAGTCTAGAGGA 4799  
QY 4681 ACCCTTACCTGGAATCTGGAATCAGCCTCTTCTGATGACCTGAATCTGATCTCT 4740  
DB 4800 ACCCTTACCTGGAATCTGGAATCAGCCTCTTCTGATGACCTGAATCTGATCTCTCT 4859  
QY 4741 GAAGACAGAGCCCGAGATCAGTCTGTTGGCAATACATCTTCAACTCTGATTTG 4800  
DB 4860 GAAGACAGAGCCCGAGATCAGTCTGTTGGCAATACATCTTCAACTCTGATTTG 4919  
QY 4801 AAGTTTCCCAATTTGAAGTGTGAGAAATGCTGCAAGTCCAGTCTGCTCATCTACT 4860  
DB 4920 AAGTTTCCCAATTTGAAGTGTGAGAAATGCTGCAAGTCCAGTCTGCTCATCTACT 4979  
QY 4861 GATACGTCTGGGTATATGCAATGGAAGAAAGTGTGAGAGGAGAGGACAGAAATGACA 4920  
DB 4980 GATACGTCTGGGTATATGCAATGGAAGAAAGTGTGAGAGGAGAGGACAGAAATGACA 5039  
QY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGCTGAGCCCGAGAA 4980  
DB 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGCTGAGCCCGAGAA 5099  
QY 4981 GAATTTATGCTGCTGATCAAGTGTGCAAGAAACACACATCATCTTAATCTAAT 5040  
DB 5100 GAATTTATGCTGCTGATCAAGTGTGCAAGAAACACACATCATCTTAATCTAAT 5159  
QY 5041 ACTGAAGAGACTATCATCTTGTATGAAGAAACAGATGCTGAGTTGTGTGAAGCAGCA 5100  
DB 5160 ACTGAAGAGACTATCATCTTGTATGAAGAAACAGATGCTGAGTTGTGTGTGAAGCAGCA 5219  
QY 5101 CTGAATTAATTTCTAGGAATGCGGAGGAAATGGGTAGTTAGTATTTCTGGTGACC 5160  
DB 5220 CTGAATTAATTTCTAGGAATGCGGAGGAAATGGGTAGTTAGTATTTCTGGTGACC 5279  
QY 5161 CAGTCTATTAAGAAAGAAATGCGTGAATGACATGATTTGAAGTCTGAGAGAGATG 5220  
DB 5280 CAGTCTATTAAGAAAGAAATGCGTGAATGACATGATTTGAAGTCTGAGAGAGATG 5339  
QY 5221 GTCAATGAAGAAACCAAGGTCCTCAAGAGCAGAGAAATCCAGAGACAGAAAGATC 5280  
DB 5340 GTCAATGAAGAAACCAAGGTCCTCAAGAGCAGAGAAATCCAGAGACAGAAAGATC 5399  
QY 5281 TTGAGGGGCTAGAAATCTGTTGCTATGGGCCCTTACCAACATGCCCACAGATCACTG 5340

Db 5400 TTCAGGGGGCTAGAAATCTGTGCTATGCGCCCTTCACCAACATGCCACAGATCAACTG 5459  
OY 5341 GAATGATGCTGACAGCTGTGTGCTCTGCTGCTGAGAGAGCTTTCATCATTCACCCCTT 5400  
Db 5460 GAATGATGCTGACAGCTGTGTGCTCTGCTGCTGAGAGAGCTTTCATCATTCACCCCTT 5519  
OY 5401 GGCACAGTGTCCACCAATTTGTTGTGTGACAGCAGATGCCGACAGAGCAATATGCG 5460  
Db 5520 GGCACAGTGTCCACCAATTTGTTGTGTGACAGCAGATGCCGACAGAGCAATATGCG 5579  
OY 5461 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTGACCCGAGAGTGGTGTGGAC 5520  
Db 5580 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGTGACCCGAGAGTGGTGTGGAC 5639  
OY 5521 AGTGTAGCACTGTACAGAGTCCAGAGAGCTGACACCTGATACCCAGATCCCCAC 5580  
Db 5640 AGTGTAGCACTGTACAGAGTCCAGAGAGCTGACACCTGATACCCAGATCCCCAC 5699  
OY 5581 AGCCACTAC 5589  
Db 5700 AGCCACTAC 5708

RESULT 7  
AAV46458 standard; cDNA: 5711 BP.  
AAV46458;  
18-NOV-1998 (first entry)  
Human BRCA1 om12 polymorphism #1 cDNA.  
BRCA1: om12: human: breast and ovarian cancer predisposing gene;  
polymorphism: susceptibility; anti-oncogene; tumour suppressor;  
chromosome 17q: ss.  
Homo sapiens.  
Location/Qualifiers  
Key CDS 120..5711  
FT CDS /tag= "BRCA1 om12 protein"  
FT variation /product= "BRCA1 om12 protein"  
FT /tag= b  
FT /note= "This polymorphic variation can be a C or T  
nucleotide"

US5750400-A.  
12-MAY-1998.  
12-FEB-1997; 97US-0798691.  
12-FEB-1996; 96US-0598591.  
12-FEB-1997; 97US-0798691.  
(ONCO-) ONCORMED INC.  
Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
Scheiter DB, Zeng B;  
WPI: 1998-296774/26.  
BRCA1 om1 gene coding sequences - useful for distinguishing between  
polymorphisms and mutation(s) in the screening for disposition to  
breast or ovarian cancer  
Claim 2e; Page -; 54pp; English.  
This sequence encodes a human BRCA1 (breast and ovarian cancer  
predisposing gene) om12 gene in which a polymorphic variation occurs at  
nucleotide 2201. This sequence and other polymorphic variations of this

CC sequence are useful for the identification of an individual who may or  
CC may not have an increased susceptibility to breast or ovarian cancer.  
CC The sequences used identify gene changes which are due to polymorphisms,  
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
CC suppressor) which is involved in genetic inheritance of cancers.  
CC especially breast and ovarian cancer. It is found at human chromosome  
CC 17q which is known to be linked to cancer susceptibility, especially  
CC breast cancer. Cells containing a mutation in this gene lose the  
CC wild-type function of BRCA1 and are more susceptible to cancers.  
CC NOTE: This sequence does not appear in the specification but has been  
CC created from the wild type BRCA1 om12 gene represented in AAV46449.  
XX  
SQ Sequence 5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;  
Query Match 100.0%; Score 5587; DB 19; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 ATGGATTTATCTGCTCTTCCGCTTGAAGATACAAATGTCATTAATGTCAGAGAA 60  
Db 120 ATGGATTTATCTGCTCTTCCGCTTGAAGATACAAATGTCATTAATGTCAGAGAA 179  
OY 61 ATCTTAGAGTGTCCATCTGTCTGAGATTGATCAAGAACTGTCTCCAAAGTGTGAC 120  
Db 180 ATCTTAGAGTGTCCATCTGTCTGAGATTGATCAAGAACTGTCTCCAAAGTGTGAC 239  
OY 121 CACATATTTGCAATTTTGCATGCTGCAACCTTCAACCCAGAGAGAGGCTTCACAG 180  
Db 240 CACATATTTGCAATTTTGCATGCTGCAACCTTCAACCCAGAGAGAGGCTTCACAG 299  
OY 181 TGTCTTTATGTAAGATATATTAACCAAAAGAGCTCTACAAAGATGAGATTAGT 240  
Db 300 TGTCTTTATGTAAGATATATTAACCAAAAGAGCTCTACAAAGATGAGATTAGT 359  
OY 241 CAAGTTGTGAAGAGCTATTTGAATATCTTTGCTTTCAGCTTGACAGCTTTGGAG 300  
Db 360 CAAGTTGTGAAGAGCTATTTGAATATCTTTGCTTTCAGCTTGACAGCTTTGGAG 419  
OY 301 TATGCAACAGCTATATATTTGCAAAAGAAATTAAGTCTCCGACATCTTAAAGAT 360  
Db 420 TATGCAACAGCTATATATTTGCAAAAGAAATTAAGTCTCCGACATCTTAAAGAT 479  
OY 480 GAAGTTTATCATCAAGATATGGGCTACAAACCCGTGCCAAAGACTTCTACAGAT 539  
Db 421 GAAGTTTATCATCAAGATATGGGCTACAAACCCGTGCCAAAGACTTCTACAGAT 480  
OY 540 GAAGTTTATCATCAAGATATGGGCTACAAACCCGTGCCAAAGACTTCTACAGAT 599  
Db 481 ACTGTGAGACTCTGAGCAAGAGCAGGATATCAACCTCAAAAGAGCTGTCTACATT 540  
OY 600 ACTGTGAGACTCTGAGCAAGAGCAGGATATCAACCTCAAAAGAGCTGTCTACATT 659  
Db 541 GAATGGAGTGTGATCTCTGAGATACCGTTAATAAGCACTTATGAGTGTGGA 600  
OY 660 GAATGGAGTGTGATCTCTGAGATACCGTTAATAAGCACTTATGAGTGTGGA 719  
Db 601 GATCAAGATTTGTACAAATCAACCCCTCAGAGACCGAGGATGAATATCATGTTGATCT 660  
OY 720 GATCAAGATTTGTACAAATCAACCCCTCAGAGACCGAGGATGAATATCATGTTGATCT 779  
Db 661 GCAAAAAAGCGTGTGCAATTTCTGAGAGGATATCAAAATATGAGATCATCA 720  
OY 780 GCAAAAAAGCGTGTGCAATTTCTGAGAGGATATCAAAATATGAGATCATCA 839  
Db 721 CCCAGTAATATGATTTGACACCACTGAGAGGCTGAGAGGATCCGAGAAAG 780  
OY 840 CCCAGTAATATGATTTGACACCACTGAGAGGCTGAGAGGATCCGAGAAAG 899  
Db 781 TATCAGGAGTGTCTTTCAAACTTGATGTGAGGATGTGGACCAAAATATCATGCC 840  
OY 900 TATCAGGAGTGTCTTTCAAACTTGATGTGAGGATGTGGACCAAAATATCATGCC 959  
Db

QY	841	AGCTCATTTACAGCATGAGAACGAGATTATTTACTCACTAAAGCACAATGATGTGAA	900
Db	960	AGCTCATTTACAGCATGAGAACGAGATTATTTACTCACTAAAGCACAATGATGTGAA	1019
QY	901	AAGCTCAATTTCTTAATTAAGCAAAACAGCTCGCTTAGCAAGAGCCAACTAACAGA	960
Db	1020	AAGCTCAATTTCTTAATTAAGCAAAACAGCTCGCTTAGCAAGAGCCAACTAACAGA	1079
QY	961	TGGCTGGAAGTAAGAAACATGTAATGATAGCGGACCTCCAGCACAGAAAAAGGTA	1020
Db	1080	TGGCTGGAAGTAAGAAACATGTAATGATAGCGGACCTCCAGCACAGAAAAAGGTA	1139
QY	1021	GATGTGAATGCTGATCCCTGCTGAGAGAAACAAATGAGATTAGCAAAATCCCATGC	1080
Db	1140	GATGTGAATGCTGATCCCTGCTGAGAGAAACAAATGAGATTAGCAAAATCCCATGC	1139
QY	1081	TCAGAGAAATCCTAGAGATACTGAAGATGTTCCCTGGATTAACCTAATATAGCAGATTAC	1140
Db	1200	TCAGAGAAATCCTAGAGATACTGAAGATGTTCCCTGGATTAACCTAATATAGCAGATTAC	1259
QY	1141	AAAGTTAATGCTGCTTTCCAGAGATGATCACTGTTAGGTTCTGATGATCACAATGAT	1200
Db	1260	AAAGTTAATGCTGCTTTCCAGAGATGATCACTGTTAGGTTCTGATGATCACAATGAT	1319
QY	1201	GGGAGTCTGATCAAAATGCCAAAGTAGCTGATATTGACAGCTTCAATAGAGTAGAT	1260
Db	1320	GGGAGTCTGATCAAAATGCCAAAGTAGCTGATATTGACAGCTTCAATAGAGTAGAT	1379
QY	1261	GAAATATTTCTGCTTTCTTACAGAGAAATATGACTTACTGGCCAGTGATCCTATAGAGCTTTA	1320
Db	1380	GAAATATTTCTGCTTTCTTACAGAGAAATATGACTTACTGGCCAGTGATCCTATAGAGCTTTA	1439
QY	1321	ATATGTAAAGTGAAGAGTCACTCCAAATCACTAGAGATATATTGAAGCAAAATA	1380
Db	1440	ATATGTAAAGTGAAGAGTCACTCCAAATCACTAGAGATATATTGAAGCAAAATA	1499
QY	1381	TTTTGGAAAAACCTATTCGGAGAAAGGCAAGGCTCCCAACTTAAGCCATGTAACTGAATAAT	1440
Db	1500	TTTTGGAAAAACCTATTCGGAGAAAGGCAAGGCTCCCAACTTAAGCCATGTAACTGAATAAT	1559
QY	1441	CTAATTATAGAGACTTTGTTACTAGAGCCACACATAATATACAAAGCGTCCCTCACAAT	1500
Db	1560	CTAATTATAGAGACTTTGTTACTAGAGCCACACATAATATACAAAGCGTCCCTCACAAT	1619
QY	1501	AAATTAAGCGTAAAGAGAGACCTACATCAGGCGCTTATCCTGAGATTTTATCAATAA	1560
Db	1620	AAATTAAGCGTAAAGAGAGACCTACATCAGGCGCTTATCCTGAGATTTTATCAATAA	1679
QY	1561	GCAGATTTGGCAGTTCAAAGAGCTCTGAAATGATAAATCAGGGAGCTAAACCAACGGAG	1620
Db	1680	GCAGATTTGGCAGTTCAAAGAGCTCTGAAATGATAAATCAGGGAGCTAAACCAACGGAG	1739
QY	1621	CAGAAATGCTCAAGGATGATATTTACTATATAGTGGCTATAGAGATTAACAAAAAGTGAT	1680
Db	1740	CAGAAATGCTCAAGGATGATATTTACTATATAGTGGCTATAGAGATTAACAAAAAGTGAT	1799
QY	1681	TCTATTGAGATGAGAAAAATCCTAACCCCAATAGAAATCACTCGAAAAAGATCGCTTTC	1740
Db	1800	TCTATTGAGATGAGAAAAATCCTAACCCCAATAGAAATCACTCGAAAAAGATCGCTTTC	1859
QY	1741	AAAAAGAAACCTGAACCTATTAAGAGAGATATTAAGCAATATGGAATCGGAATTTAATATC	1800
Db	1860	AAAAAGAAACCTGAACCTATTAAGAGAGATATTAAGCAATATGGAATCGGAATTTAATATC	1919
QY	1801	CACAAATTCAAAGACCTAAAGAAAGATAGGCTGAGAGAAAGTCTTCTACAGGCAATTT	1860
Db	1920	CACAAATTCAAAGACCTAAAGAAAGATAGGCTGAGAGAAAGTCTTCTACAGGCAATTT	1979
QY	1861	CATCGCGTTGACATAGTACTAGTAAGATCTAAGCCCACTAATTTACTGAATTCAA	1920
Db	1980	CATCGCGTTGACATAGTACTAGTAAGATCTAAGCCCACTAATTTACTGAATTCAA	2039

QY	1921	ATTGTA	TACTTGTCT	TACAGAGTGAAGATTA	AGAAAAA	AAAGTCA	AAATCCAA	ATCCGAC	TC	1980		
Db	2040	ATTGTA	TACTTGTCT	TACAGAGTGAAGATTA	AGAAAAA	AAAGTCA	AAATCCAA	ATCCGAC	TC	2099		
QY	1981	AGGCAC	CACACAAACCT	CAACCTCAT	TGGAGAGTAA	AGAACCT	CGACCTG	AGCCACA	AG	2040		
Db	2100	AGGCAC	CACACAAACCT	CAACCTCAT	TGGAGAGTAA	AGAACCT	CGACCTG	AGCCACA	AG	2159		
QY	2041	AGTAACA	AGCCAAATG	AAACAGACAGATTA	AAAGACATG	ACATGAT	TACTTTCC	GAGAC	TCG	2100		
Db	2160	AGTAACA	AGCCAAATG	AAACAGACAGATTA	AAAGACATG	ACATGAT	TACTTTCC	GAGAC	TCG	2219		
QY	2101	AAGTTA	ACAATTCG	ACCTGGTCTT	TTTACTTA	AGTGTCCA	ANTACCA	GTGAAC	TTAAAGAA	2160		
Db	2220	AAGTTA	ACAATTCG	ACCTGGTCTT	TTTACTTA	AGTGTCCA	ANTACCA	GTGAAC	TTAAAGAA	2279		
QY	2161	TTTGCA	TCTCTAG	CCCTCCCA	AGAGAA	AAAAAACA	AGAACTAG	AAACAG	TTAAAGTG	2220		
Db	2280	TTTGCA	TCTCTAG	CCCTCCCA	AGAGAA	AAAAAACA	AGAACTAG	AAACAG	TTAAAGTG	2339		
QY	2221	TCTAATA	TGCTGA	AGACCCCA	AGAGATCTCA	TGTTAAGT	GGAGAA	GGGTTT	GGCAACT	2280		
Db	2340	TCTAATA	TGCTGA	AGACCCCA	AGAGATCTCA	TGTTAAGT	GGAGAA	GGGTTT	GGCAACT	2399		
QY	2281	GAAGAT	CTGTAG	AGAGTAC	GAATTTCA	TGTTACCT	GGTACTG	TTATG	CAC	TCAG	2340	
Db	2400	GAAGAT	CTGTAG	AGAGTAC	GAATTTCA	TGTTACCT	GGTACTG	TTATG	CAC	TCAG	2459	
QY	2341	GAAGAT	CTGTAG	AGAGTAC	GAATTTCA	TGTTACCT	GGTACTG	TTATG	CAC	TCAG	2400	
Db	2460	GAAGAT	CTGTAG	AGAGTAC	GAATTTCA	TGTTACCT	GGTACTG	TTATG	CAC	TCAG	2519	
QY	2401	TGTGTG	ATCTAG	TCTG	CGACATTTT	GAAACCCCA	AGAGACTA	ATTCAT	TGTTTCC	CAAA	2460	
Db	2520	TGTGTG	ATCTAG	TCTG	CGACATTTT	GAAACCCCA	AGAGACTA	ATTCAT	TGTTTCC	CAAA	2579	
QY	2461	GATAAT	AGAAATG	ACACAGA	AGCTTT	PAGTATTC	ATTCAT	TGGACAT	GATGA	TTAAC	CACAGT	2520
Db	2580	GATAAT	AGAAATG	ACACAGA	AGCTTT	PAGTATTC	ATTCAT	TGGACAT	GATGA	TTAAC	CACAGT	2639
QY	2521	CGGGA	ACAACCA	CTATGA	AAATGAGAA	AGAACTG	ATGCTG	CACTTTT	GGCAAG	ATACA	2580	
Db	2640	CGGGA	ACAACCA	CTATGA	AAATGAGAA	AGAACTG	ATGCTG	CACTTTT	GGCAAG	ATACA	2699	
QY	2581	TTTCA	AGTTTAA	AGCGC	AGCTCAT	TTTGTG	TTTCCAAT	CAGAGAA	TG	CAGAGAG	2640	
Db	2700	TTTCA	AGTTTAA	AGCGC	AGCTCAT	TTTGTG	TTTCCAAT	CAGAGAA	TG	CAGAGAG	2759	
QY	2641	GAATGT	GCACAT	TCTGTG	CCCACTGT	GGTCTT	TAAGAA	CAAAAGT	CC	CAAAAGT	CAC	2700
Db	2760	GAATGT	GCACAT	TCTGTG	CCCACTGT	GGTCTT	TAAGAA	CAAAAGT	CC	CAAAAGT	CAC	2819
QY	2701	TTTGA	ATCTGA	CAAAAGS	AGAAATACA	AGSAAATG	AGTCTA	ATATCA	AC	CCGTGA	2760	
Db	2820	TTTGA	ATCTGA	CAAAAGS	AGAAATACA	AGSAAATG	AGTCTA	ATATCA	AC	CCGTGA	2879	
QY	2761	CAGAC	AGTTAAT	ATAC	TCAGAGCTT	CTGCTGTG	TTGGT	CAGAA	AGATTA	AGCCAG	TGTAT	2820
Db	2880	CAGAC	AGTTAAT	ATAC	TCAGAGCTT	CTGCTGTG	TTGGT	CAGAA	AGATTA	AGCCAG	TGTAT	2939
QY	2821	AATGCA	AATGTAG	TATCA	AAAGAGGCTT	AGTTTTGT	CTAT	ATCAT	CTCAG	TTTCCAG	AGGC	2880
Db	2940	AATGCA	AATGTAG	TATCA	AAAGAGGCTT	AGTTTTGT	CTAT	ATCAT	CTCAG	TTTCCAG	AGGC	2999
QY	2881	AACGA	AACTG	CATCTT	ACTCCAA	ATTAACAT	GTGACTT	TACAA	AACCCAT	ATCGTATA	2940	
Db	3000	AACGA	AACTG	CATCTT	ACTCCAA	ATTAACAT	GTGACTT	TACAA	AACCCAT	ATCGTATA	3059	
QY	2941	CCAC	CACTTTT	TTCCAT	CAACGATCT	TTTGGTT	TTAAAC	TAAATGT	TAAGAA	AAATCTG	CTGAG	3000
Db	3060	CCAC	CACTTTT	TTCCAT	CAACGATCT	TTTGGTT	TTAAAC	TAAATGT	TAAGAA	AAATCTG	CTGAG	3119
QY	3001	GAAAC	ATTTG	AGACAT	TCATGAT	CTGCAC	CTGAA	AGAGAAAT	TGGAAAT	TGAAC	ATTTCCA	3060



Db 3120 GAAACATTTGAGGAGACATTCATGATCACCCTGAAGAGAAATGGAAATGAGAACATTCACA 3179  
Qy 3061 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAAATGTTTTTAAGAACCCAGC 3120  
Db 3180 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAAATGTTTTTAAGAACCCAGC 3239  
Qy 3121 TCAGCAATATTAATGAAGTGGTCCAGTACTAATGAATGGGCTCCAGTATTAATGA 3180  
Db 3240 TCAGCAATATTAATGAAGTGGTCCAGTACTAATGAATGGGCTCCAGTATTAATGA 3299  
Qy 3181 ATAGGTTCCAGTATGAAAAATTCAGACGAACTAGGTAGAAACAGAGGCCAAAAATG 3240  
Db 3300 ATAGGTTCCAGTATGAAAAATTCAGACGAACTAGGTAGAAACAGAGGCCAAAAATG 3359  
Qy 3241 AATGCTATGCTTATGATTTAGGGTTTTGCAACCTGAGTCTATTAACAAATCTTCCGGA 3300  
Db 3360 AATGCTATGCTTATGATTTAGGGTTTTGCAACCTGAGTCTATTAACAAATCTTCCGGA 3419  
Qy 3301 AGTAATGTAGCATCTGTAATAATAAAGCAAGAAATATGAAGAGTTCAGACTGT 3360  
Db 3420 AGTAATGTAGCATCTGTAATAATAAAGCAAGAAATATGAAGAGTTCAGACTGT 3479  
Qy 3361 AATPACGATTTCTCTCCATATCTGATTTAGATTAAGACAGCCCTATGGAGTAGT 3420  
Db 3480 AATPACGATTTCTCTCCATATCTGATTTAGATTAAGACAGCCCTATGGAGTAGT 3539  
Qy 3421 CATGCATCTCAGGTTTTGTTCTGAGACACCGATGACCTGTATGATGATGGTGAATTAAG 3480  
Db 3540 CATGCATCTCAGGTTTTGTTCTGAGACACCGATGACCTGTATGATGATGGTGAATTAAG 3599  
Qy 3481 GAAGATACTAGTTTGTGTAATAATGACATTAAGAAAGTTCTGCTGTTTTTACCAAAAGC 3540  
Db 3600 GAAGATACTAGTTTGTGTAATAATGACATTAAGAAAGTTCTGCTGTTTTTACCAAAAGC 3659  
Qy 3541 GTCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTTACCCATACACTTTGGCTCAG 3600  
Db 3660 GTCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTTACCCATACACTTTGGCTCAG 3719  
Qy 3601 GGTACCCGAGAGGGGCCAAGAAATTAGAGTCCCTGAGAGAGAACTTATCTAGTGAAGAT 3660  
Db 3720 GGTACCCGAGAGGGGCCAAGAAATTAGAGTCCCTGAGAGAGAACTTATCTAGTGAAGAT 3779  
Qy 3661 GAAGAGCTTCCCTGCTTCCACACATTTGTTTGAAGTAAACAAATATACCTTCTCAG 3720  
Db 3780 GAAGAGCTTCCCTGCTTCCACACATTTGTTTGAAGTAAACAAATATACCTTCTCAG 3839  
Qy 3721 TCTACTAGCAATGACCCGTTGCTACCCGAGTGTCTGTCTTAAGAACACAGAGAGAAATTA 3780  
Db 3840 TCTACTAGCAATGACCCGTTGCTACCCGAGTGTCTGTCTTAAGAACACAGAGAGAAATTA 3899  
Qy 3781 TTATCATTTGAAGAAATAGCTTAAATGACTGACGTAACAGAGTAATTTGGCAAGGCATCT 3840  
Db 3900 TTATCATTTGAAGAAATAGCTTAAATGACTGACGTAACAGAGTAATTTGGCAAGGCATCT 3959  
Qy 3841 CAGGAACATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTGTCTTCCACAGTGC 3900  
Db 3960 CAGGAACATCACCTTAGTAGAGAAACAAATGTTCTGCTAGCTTGTCTTCCACAGTGC 4019  
Qy 3901 AGTGAATTTGAAGACTTGAATGCTCAAAATACAAACACCCAGAGATCCTTCTTGATTTGTTCT 3960  
Db 4020 AGTGAATTTGAAGACTTGAATGCTCAAAATACAAACACCCAGAGATCCTTCTTGATTTGTTCT 4079  
Qy 3961 TCCAAACAATAGGATCAGTCTGAAAGCCAGAGGATTTGTTGATGAGTGAAGGAATG 4020  
Db 4080 TCCAAACAATAGGATCAGTCTGAAAGCCAGAGGATTTGTTGATGAGTGAAGGAATG 4139  
Qy 4021 GTTTCAGATGATGAAGAAAGAGAGAGGCTTTGGAAGAAATATATCAAGAGCAAGC 4080  
Db 4140 GTTTCAGATGATGAAGAAAGAGAGAGGCTTTGGAAGAAATATATCAAGAGCAAGC 4199  
Qy 4081 ATGGATTCAAACTTAGGTAGAGCAGCATCTGGGTGTGAGAGTGAACAAGCCTCTCTGAA 4140  
Db 4200 ATGGATTCAAACTTAGGTAGAGCAGCATCTGGGTGTGAGAGTGAACAAGCCTCTCTGAA 4259  
Qy 4141 GACTGCTCAGGGGCTATCTCTGAGTGAACATTTTAACCACTCAGCAGAGGATCCATG 4200  
Db 4260 GACTGCTCAGGGGCTATCTCTGAGTGAACATTTTAACCACTCAGCAGAGGATCCATG 4319  
Qy 4201 CAACATTAACCTGATTAAGCTCCAGCAGGAAATGGCTGAACCTAGAAGCTGTGTTAGAACAG 4260  
Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGGAAATGGCTGAACCTAGAAGCTGTGTTAGAACAG 4379  
Qy 4261 CATGGAGCCAGGCTTCTTAACAGCTACCTCCATCATATAGTACTCTTGTGCCCTTGAG 4320  
Db 4380 CATGGAGCCAGGCTTCTTAACAGCTACCTCCATCATATAGTACTCTTGTGCCCTTGAG 4439  
Qy 4321 GACCTGCGAAATCCAGAACAGACATGAGAAAAAGCATTTTACTTCCACAGAAAGT 4380  
Db 4440 GACCTGCGAAATCCAGAACAGACATGAGAAAAAGCATTTTACTTCCACAGAAAGT 4499  
Qy 4381 AGTGAATACCTTATTAAGCCAGATCCAGAGGCTTCTGCTGACAAAGTTTGAAGTCTCT 4440  
Db 4500 AGTGAATACCTTATTAAGCCAGATCCAGAGGCTTCTGCTGACAAAGTTTGAAGTCTCT 4559  
Qy 4441 GCAGATAGTTCTTACAGTAAAAATTAAGAACAGAGAGTGAAGAGTCAATCCCTTCTTAA 4500  
Db 4560 GCAGATAGTTCTTACAGTAAAAATTAAGAACAGAGAGTGAAGAGTCAATCCCTTCTTAA 4619  
Qy 4501 TGCCCATCATTTAGATAGTGGTGTGATGATGACAGTGTGCTGGAGTCTCAGAAATGA 4560  
Db 4620 TGCCCATCATTTAGATAGTGGTGTGATGATGACAGTGTGCTGGAGTCTCAGAAATGA 4679  
Qy 4561 AACTACCATCTTCAAGAGAGGCTCATTAAGTTGTTGATGATGAGAGACCAACAGCTGGA 4620  
Db 4680 AACTACCATCTTCAAGAGAGGCTCATTAAGTTGTTGATGATGAGAGACCAACAGCTGGA 4739  
Qy 4621 GAGTCTGGGCGCACACGATTTGACGGGAAACATCTTACTTGTCCAGAGCAGATCTAGAGGA 4680  
Db 4740 GAGTCTGGGCGCACACGATTTGACGGGAAACATCTTACTTGTCCAGAGCAGATCTAGAGGA 4799  
Qy 4681 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTCTGATGAGCCCTGGAATCTGATCTCT 4740  
Db 4800 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTCTGATGAGCCCTGGAATCTGATCTCT 4859  
Qy 4741 GAAGACAGAGCCCCAGAGTCAAGTGTGTCACACATACATCTTCAACCTCTGCAATG 4800  
Db 4860 GAAGACAGAGCCCCAGAGTCAAGTGTGTCACACATACATCTTCAACCTCTGCAATG 4919  
Qy 4801 AAGGTTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTGCTATACTACT 4860  
Db 4920 AAGGTTCCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTGCTATACTACT 4979  
Qy 4861 GATACGCTGGGTATTAATGCAATGAGAGAAAGTGTGACAGAGGAGAAACCGAAATGACA 4920  
Db 4980 GATACGCTGGGTATTAATGCAATGAGAGAAAGTGTGACAGAGGAGAAACCGAAATGACA 5039  
Qy 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCTCAGATGCTGCTGACCCAGAA 4980  
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCTCAGATGCTGCTGACCCAGAA 5099  
Qy 4981 GAATTTATCTGCTGTAGCAAGTTTCCAGAAAAACCAATCACTTAACTAAATCTAAT 5040  
Db 5100 GAATTTATCTGCTGTAGCAAGTTTCCAGAAAAACCAATCACTTAACTAAATCTAAT 5159  
Qy 5041 ACTGAAGACATCTATCATGTTGTTATGAAGACAGATGCTGAGTTTGTGTGAACGGACA 5100  
Db 5160 ACTGAAGACATCTATCATGTTGTTATGAAGACAGATGCTGAGTTTGTGTGAACGGACA 5219  
Qy 5101 CTGAATATATTTCTAGGAATTTGCGGAGAGAAATGGATAGTAACTATTTCTGGGTACC 5160  
Db 5220 CTGAATATATTTCTAGGAATTTGCGGAGAGAAATGGATAGTAACTATTTCTGGGTACC 5279  
Qy 5161 CAGTCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGTG 5220  
Db 5280 CAGTCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGTG 5339



QY 5221 GTCATGGAAGAAACACCAAGGTCCAAAGCGAGAGAGATCCAGACAGAAAGATC 5280  
 |||||||  
 Db 5340 GTCATGGAAGAAACACCAAGGTCCAAAGCGAGAGAGATCCAGACAGAAAGATC 5339  
 QY 5281 TTCAGGGGGGTAGAAATCTGTGTATGAGGCCCTTCCACCAACATGCCCCAGATCAACTG 5340  
 |||||||  
 Db 5400 TTCAGGGGGGTAGAAATCTGTGTATGAGGCCCTTCCACCAACATGCCCCAGATCAACTG 5459  
 QY 5341 GAATGATGGGTACAGCTGTGTGTCTTCTGTGTGAAGAGCTTCATCATTCACCTT 5400  
 |||||||  
 Db 5460 GAATGATGGGTACAGCTGTGTGTCTTCTGTGTGAAGAGCTTCATCATTCACCTT 5519  
 QY 5401 GGCACAGGTGTCCACCAATTTGTGTGTGACAGCCAGATGCTTGACAGAGACAAATGGC 5460  
 |||||||  
 Db 5520 GGCACAGGTGTCCACCAATTTGTGTGTGACAGCCAGATGCTTGACAGAGACAAATGGC 5579  
 QY 5461 TTCATGCAATTTGGGACAGATGTGAGGACCTTGTGTGACCCAGAGTGGGTGTGAC 5520  
 |||||||  
 Db 5580 TTCATGCAATTTGGGACAGATGTGAGGACCTTGTGTGACCCAGAGTGGGTGTGAC 5639  
 QY 5521 AGTGTACACTCTACAGTGTGACAGAGTGTGACACCTACCTGATACCCAGATCCCCAC 5580  
 |||||||  
 Db 5640 AGTGTACACTCTACAGTGTGACAGAGTGTGACACCTACCTGATACCCAGATCCCCAC 5699  
 QY 5581 AGCCACTAC 5589  
 |||||||  
 Db 5700 AGCCACTAC 5708  
 RESULT 8  
 AAV46459  
 ID AAV46459 standard; cDNA; 5711 BP.  
 AC AAV46459;  
 XX  
 DT 18-NOV-1998 (first entry)  
 XX  
 DE Human BRCA1 omi2 polymorphism #2 cDNA.  
 XX  
 KW BRCA1: omi2: human: breast and ovarian cancer predisposing gene:  
 KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;  
 KW chromosome 17q; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 120..5711  
 FT /\*tag= a  
 FT /product= "BRCA1 omi2 protein"  
 FT variation 2430  
 FT /\*tag= b  
 FT /note= "this polymorphic variation can be a T or C  
 FT nucleotide"  
 FT  
 XX US5750400-A.  
 XX PN 12-MAY-1998.  
 XX PD 12-FEB-1997; 97US-0798691.  
 XX FE 12-FEB-1997; 97US-0798691.  
 XX PR 12-FEB-1996; 96US-0598591.  
 XX PR 12-FEB-1997; 97US-0798691.  
 XX PA (ONCO-) ONCOMED INC.  
 XX PI Allen AC, Alvarez CP, Critz BS, Murphy PD, Olson SJ;  
 XX PI Schelter DB, Zeng B;  
 XX DR WPI: 1998-296774/26.  
 XX PT BRCA1 omi gene coding sequences - useful for distinguishing between  
 PT polymorphisms and mutation(s) in the screening for disposition to

PT breast or ovarian cancer  
 XX  
 PS Claim 2e: Page -: 54pp: English.  
 XX  
 CC This sequence encodes a human BRCA1 (breast and ovarian cancer  
 CC -predisposing gene) omi2 gene in which a polymorphic variation occurs at  
 CC nucleotide 2430. This sequence and other polymorphic variations of this  
 CC sequence are useful for the identification of an individual who may or  
 CC may not have an increased susceptibility to breast or ovarian cancer.  
 CC The sequences used identify gene changes which are due to polymorphisms,  
 CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour  
 CC suppressor) which is involved in genetic inheritance of cancers,  
 CC especially breast and ovarian cancer. It is found at human chromosome  
 CC 17q which is known to be linked to cancer susceptibility, especially  
 CC breast cancer. Cells containing a mutation in this gene lose the  
 CC wild-type function of BRCA1 and are more susceptible to cancers.  
 CC NOTE: This sequence does not appear in the specification but has been  
 CC created from the wild type BRCA1 omi2 gene represented in AAV46449.  
 CC  
 SQ Sequence 5711 BP: 1956 A; 1098 C; 1274 G; 1382 T; 1 other;  
 Query Match 100.0%; Score 5587; DB 19; Length 5711;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5587; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGATTTATCTGCTTCGCGTTGAAGAAAGTACAAATGCTTAATGCTATGCAGAA 60  
 |||||||  
 Db 120 ATGATTTATCTGCTTCGCGTTGAAGAAAGTACAAATGCTTAATGCTATGCAGAA 179  
 QY 61 ATCTTAAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCCCAAAAGTGTAC 120  
 |||||||  
 Db 180 ATCTTAAGTGTCCATCTGTCTGAGTGTATCAAGAAACCTGTCCCAAAAGTGTAC 239  
 QY 121 CACATATTTGCCAATTTTGCATGCTGAACCTTCTCAACCGAAGAAAGGCCCTTCACAG 180  
 |||||||  
 Db 240 CACATATTTGCCAATTTTGCATGCTGAACCTTCTCAACCGAAGAAAGGCCCTTCACAG 299  
 QY 181 TGTCTTTATGTAGAAATGATATTAACCAAAAGAGCCTACAGAAAGTACAGATTTAGT 240  
 |||||||  
 Db 300 TGTCTTTATGTAGAAATGATATTAACCAAAAGAGCCTACAGAAAGTACAGATTTAGT 359  
 QY 241 CAACCTGTGTAGAGCTATTTGAANAATCTTGTCTTTTTCAGCTTGACACAGTTTGGAG 300  
 |||||||  
 Db 360 CAACCTGTGTAGAGCTATTTGAANAATCTTGTCTTTTTCAGCTTGACACAGTTTGGAG 419  
 QY 301 TATGCAAAACAGCTATTAATTTTGCAGAAAGAAATTAATCTCTGAAACATCTAAAGAT 360  
 |||||||  
 Db 420 TATGCAAAACAGCTATTAATTTTGCAGAAAGAAATTAATCTCTGAAACATCTAAAGAT 479  
 QY 361 GAATTTCTATCATCCAAAGATGAGCTACAGAAAGCGTGCAAAAGACTCTACAGAGT 420  
 |||||||  
 Db 480 GAATTTCTATCATCCAAAGATGAGCTACAGAAAGCGTGCAAAAGACTCTACAGAGT 539  
 QY 421 GAACCCGAAAAATCCTTCCTGCGAGAAACAGTCTCAGTGTCAACTCTCTAACCTTGA 480  
 |||||||  
 Db 540 GAACCCGAAAAATCCTTCCTGCGAGAAACAGTCTCAGTGTCAACTCTCTAACCTTGA 599  
 QY 481 ACTGTGAACCTGTGAGAGCAACAGCGGATACAACTCAAAAGAGCTCTGTACATT 540  
 |||||||  
 Db 600 ACTGTGAACCTGTGAGAGCAACAGCGGATACAACTCAAAAGAGCTCTGTACATT 659  
 QY 541 GAATTTGGATCTGATCTTCTGGAAGATACCGTTTAATAGGCAACTTATTTGCAAGTGGGA 600  
 |||||||  
 Db 660 GAATTTGGATCTGATCTTCTGGAAGATACCGTTTAATAGGCAACTTATTTGCAAGTGGGA 719  
 QY 601 GATCAGAATTTGTTCAAAATCACCCCTCAAGGAACAGAGGATGAATCAGTTTGGATTCT 660  
 |||||||  
 Db 720 GATCAGAATTTGTTCAAAATCACCCCTCAAGGAACAGAGGATGAATCAGTTTGGATTCT 779  
 QY 661 GCAAAAAAGGCTGCTGTGGAATTTTTCGAGAGCGATTAACAAATACTGAACATCATCAA 720  
 |||||||  
 Db 780 GCAAAAAAGGCTGCTGTGGAATTTTTCGAGAGCGATTAACAAATACTGAACATCATCAA 839

QY 721 CCCAGTAATGATTTTGAACACACCTGAGAACGCTGACCTGAGAGCATCCAGAAAG 780  
|||||  
Db 840 CCCAGTAATGATTTTGAACACACCTGAGAACGCTGAGAGCATCCAGAAAG 899  
QY 781 TATCAGGGTAGTTCCTTTCAAACTTGACGTGGAGCCATGTGGCACAATTAATCTATGCC 840  
|||||  
Db 900 TATCAGGGTAGTTCCTTTCAAACTTGACGTGGAGCCATGTGGCACAATTAATCTATGCC 959  
QY 841 ACCTCATTTACAGATGAGAACAGCAGTATTTATCTACCTAAGACAGCAATGATGTAGAA 900  
|||||  
Db 960 ACCTCATTTACAGATGAGAACAGCAGTATTTATCTACCTAAGACAGCAATGATGTAGAA 1019  
QY 901 AAGGCTGAATTCCTGTATATAAAGCAACACGCTGGCTTACCAAGAGAGCCAACTAATACAGA 960  
1020 AAGGCTGAATTCCTGTATATAAAGCAACACGCTGGCTTACCAAGAGAGCCAACTAATACAGA 1079  
QY 961 TGGGCTGGAGTAGTAGGAACACATGTATGATAGCGGAGCTCCGACACAGAAAAAGTA 1020  
|||||  
Db 1080 TGGGCTGGAGTAGTAGGAACACATGTATGATAGCGGAGCTCCGACACAGAAAAAGTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAAGTAATGAGAGAAAGTGGCATGC 1080  
|||||  
Db 1140 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAAGTAATGAGAGAAAGTGGCATGC 1199  
QY 1081 TCAGAGAAATCCTAGAGATCTGAAAGATGTTCTTGGATTAACACTAATATAGCAGCATTCAG 1140  
1200 TCAGAGAAATCCTAGAGATCTGAAAGATGTTCTTGGATTAACACTAATATAGCAGCATTCAG 1259  
QY 1141 AAGGTAATGAGTGGTTCCTCCAGAACTGATGAACTGTGAGTTCATGATCTACATGAT 1200  
|||||  
Db 1260 AAGGTAATGAGTGGTTCCTCCAGAACTGATGAACTGTGAGTTCATGATCTACATGAT 1319  
QY 1201 GGGGAGTCTGAATCAAAATGCCAAAGTAGTGTATGAGACGTTTAAATGAGTAGAT 1260  
|||||  
Db 1320 GGGGAGTCTGAATCAAAATGCCAAAGTAGTGTATGAGACGTTTAAATGAGTAGAT 1379  
QY 1261 GAATATTCCTGTTCTTCAAGAAATAGACTTAATGCGCAGATGCTCTATGAGCTTTA 1320  
|||||  
Db 1380 GAATATTCCTGTTCTTCAAGAAATAGACTTAATGCGCAGATGCTCTATGAGCTTTA 1439  
QY 1321 AATATGAAAGTGAAGAGTTCCTCAATCAGTGAAGATTAATGTAAGCAAAATA 1380  
|||||  
Db 1440 AATATGAAAGTGAAGAGTTCCTCAATCAGTGAAGATTAATGTAAGCAAAATA 1499  
QY 1381 TTTGGGAAAACCTTACGGAAGAGGCAACGCTCCCACTTAAGCCATGTAATGAAAT 1440  
1500 TTTGGGAAAACCTTACGGAAGAGGCAACGCTCCCACTTAAGCCATGTAATGAAAT 1559  
QY 1441 CTAATTTATAGAGCATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAAT 1500  
1560 CTAATTTATAGAGCATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAAT 1619  
QY 1501 AAATTAAGCCATAAAGGAGCTTCATCAGAGCCCTCATCCTGAGATTTTATCAAGAAA 1560  
|||||  
Db 1620 AAATTAAGCCATAAAGGAGCTTCATCAGAGCCCTCATCCTGAGATTTTATCAAGAAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAAGACTCCTGAATGATTAATCAGGAACTAAACAAAGGAG 1620  
1680 GCAGATTTGGCAGTTCAAAAAGACTCCTGAATGATTAATCAGGAACTAAACAAAGGAG 1739  
QY 1621 CAGAAATGCTCAAGTATGATTAATTAATAGTGGTCATGAGATTAACAAAGGATGAT 1680  
1740 CAGAAATGCTCAAGTATGATTAATTAATAGTGGTCATGAGATTAACAAAGGATGAT 1799  
QY 1681 TCTATTCAGATGAGAAAAATCCTAACCAATGATCACTCGAAAAAGAAATGCTGCTTC 1740  
|||||  
Db 1800 TCTATTCAGATGAGAAAAATCCTAACCAATGATCACTCGAAAAAGAAATGCTGCTTC 1859  
QY 1741 AAAAGGAAAGCTGAACCTTAAGACAGCAGTATTAAGCAATATGAGAACTCGAATTAATATC 1800  
1860 AAAAGGAAAGCTGAACCTTAAGACAGCAGTATTAAGCAATATGAGAACTCGAATTAATATC 1919  
QY 1801 CACAATTCAAAAAGCCTTAATAAAGATAGGCTGAGAGGAAGTCTTCTACAGCATATTT 1860  
|||||  
Db 1920 CACAATTCAAAAAGCCTTAATAAAGATAGGCTGAGAGGAAGTCTTCTACAGCATATTT 1979  
QY 1861 CATGCCCTGAACTACTACTAGTCAATCTAAGGCCCACTTAATTTGATCTGAATTTGCA 1920  
1980 CATGCCCTGAACTACTACTAGTCAATCTAAGGCCCACTTAATTTGATCTGAATTTGCA 2039  
QY 1921 ATTGATAGTGTCTGACAGTGAAGATTAAGAAAAAAGTATACCAACCAATGCGCAGTC 1980  
2040 ATTGATAGTGTCTGACAGTGAAGATTAAGAAAAAAGTATACCAACCAATGCGCAGTC 2099  
QY 1981 AGGCACAGAGAAAACCTTACACTCATGAGGTAAGAAACCTGCAACTGAGACCAAGAAG 2040  
2100 AGGCACAGAGAAAACCTTACACTCATGAGGTAAGAAACCTGCAACTGAGACCAAGAAG 2159  
QY 2041 AGTAACAGCCAAATGAACAGACAGCAATTAACACATGACAGTACTTCCAGAGCTG 2100  
2160 AGTAACAGCCAAATGAACAGACAGCAATTAACACATGACAGTACTTCCAGAGCTG 2219  
QY 2101 AAGTTAACAAATGACACCTGTTCTTTACTAAGTGTCAATTAACAGTGAACCTTAAGAA 2160  
2220 AAGTTAACAAATGACACCTGTTCTTTACTAAGTGTCAATTAACAGTGAACCTTAAGAA 2279  
QY 2161 TTTGTCAATCTAGCCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2220  
2280 TTTGTCAATCTAGCCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2339  
QY 2221 TCTAATATGCTGAACACCCCAAGATCTCATGTTAAGTGGGAAAGGTTTTGCAAACT 2280  
2340 TCTAATATGCTGAACACCCCAAGATCTCATGTTAAGTGGGAAAGGTTTTGCAAACT 2399  
QY 2281 GAAAGATCTGTAGAGTAGAGTACAGTATTTCAATTTGTAAGTCACTGATTAATGAGCAGT 2340  
2400 GAAAGATCTGTAGAGTAGAGTACAGTATTTCAATTTGTAAGTCACTGATTAATGAGCAGT 2459  
QY 2341 GAAAGATCTGTAGAGTAGAGTACAGTATTTGTAAGTCACTGATTAATGAGCAGT 2400  
2460 GAAAGATCTGTAGAGTAGAGTACAGTATTTGTAAGTCACTGATTAATGAGCAGT 2519  
QY 2401 TGTGTAGTCAAGTGTGACAGTATTTGAAACCCCAAGAGGCACTAATTCATGTTGTTCCAAA 2460  
2520 TGTGTAGTCAAGTGTGACAGTATTTGAAACCCCAAGAGGCACTAATTCATGTTGTTCCAAA 2579  
QY 2461 GATTAATGAAATGACACAGAGGCTTTAAGTATCCATTTGGACATGAATTAACACAGT 2520  
2580 GATTAATGAAATGACACAGAGGCTTTAAGTATCCATTTGGACATGAATTAACACAGT 2639  
QY 2521 CGGGAACAAAGCATTAAGAAATGGAAGAAAGTGAATGATGCTCAGTATTTGACAGATACA 2580  
2640 CGGGAACAAAGCATTAAGAAATGGAAGAAAGTGAATGATGCTCAGTATTTGACAGATACA 2699  
QY 2581 TTTCAAGTTTCAAGAGCGCCAGTATTTGCTGTTTCAAAATCCAGGAATGACAGAAAG 2640  
2700 TTTCAAGTTTCAAGAGCGCCAGTATTTGCTGTTTCAAAATCCAGGAATGACAGAAAG 2739  
QY 2641 GAATGTGCAACATTTCTGCCCACTCTGGGTCCTTTAAGAAAAAAGTCCAAAGTCACT 2700  
2760 GAATGTGCAACATTTCTGCCCACTCTGGGTCCTTTAAGAAAAAAGTCCAAAGTCACT 2819  
QY 2701 TTTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCTGTA 2760  
2820 TTTGAATGTGAACAAAGAGAAAGAAATCAAGAAAGAAATGAGTCTAATATCAAGCTGTA 2879  
QY 2761 CAGACAGTTAATATCACTGACAGCTTTCCCTGTTGGTGGCAGAAAGATTAACAGTTCAT 2820  
2880 CAGACAGTTAATATCACTGACAGCTTTCCCTGTTGGTGGCAGAAAGATTAACAGTTCAT 2939  
QY 2821 AATGCCAAATGTAGATCAAAAGAGGCTCTAGGTTTGTCTATCACTCACTTCAAGAGC 2880  
2940 AATGCCAAATGTAGATCAAAAGAGGCTCTAGGTTTGTCTATCACTCACTTCAAGAGC 2999  
QY 2881 AACGAATGTGACTCTTATCTCCAAATTAACATGACCTTTTACAAACCCATATCTATA 2940  
|||||

Db 3000 AACGAACCTGGACTATCTCCAAATTAACATGACCTTTTACAAAACCATATCGTATA 3059  
 QY 2941 CCACCACTTTTCCCATCAAGTCATTTGTTAAACATAATGTAAGAAAAATGCTAGAG 3000  
 Db 3060 CCACCACTTTTCCCATCAAGTCATTTGTTAAACATAATGTAAGAAAAATGCTAGAG 3119  
 QY 3001 GAAAACTTTGAGGAACATTCATGTCACCTGAAAGAGAAAAATGGGAAATGAGAACATTTCCA 3060  
 Db 3120 GAAAACTTTGAGGAACATTCATGTCACCTGAAAGAGAAAAATGGGAAATGAGAACATTTCCA 3179  
 QY 3061 AGTACAGTGAACACATTTACCCGTAATACATTTAGAGAAAAATGTTTAAAGAACCCAGC 3120  
 Db 3180 AGTACAGTGAACACATTTACCCGTAATACATTTAGAGAAAAATGTTTAAAGAACCCAGC 3239  
 QY 3121 TCAGCAATATTTATGAAAGTAGTTCAGTACTAATGAGTGGGCTCAGTATTAATGAA 3180  
 Db 3240 TCAGCAATATTTATGAAAGTAGTTCAGTACTAATGAGTGGGCTCAGTATTAATGAA 3299  
 QY 3181 ATAGGTTCCAGTATGAAACATTTCAAGCAAGAACTAGTAGAAAACAGAGGGCCAAAATTG 3240  
 Db 3300 ATAGGTTCCAGTATGAAACATTTCAAGCAAGAACTAGTAGAAAACAGAGGGCCAAAATTG 3359  
 QY 3241 AATGCTATGCTTATGATTTAGAGGGTTTTCACACCTGAGTCTATTAACAAAGTCTTCTGGA 3300  
 Db 3360 AATGCTATGCTTATGATTTAGAGGGTTTTCACACCTGAGTCTATTAACAAAGTCTTCTGGA 3419  
 QY 3301 AGTATATGTAAGCATCTGTAATTAAGAAACAGAAATATGAGAGTATGAGTCTGAGACTGT 3360  
 Db 3420 AGTATATGTAAGCATCTGTAATTAAGAAACAGAAATATGAGAGTATGAGTCTGAGACTGT 3479  
 QY 3361 AATACAGATTTCTCTCCATATCTGATTTAGATTAAGTACAGACCTATGAGGAGTATG 3420  
 Db 3480 AATACAGATTTCTCTCCATATCTGATTTAGATTAAGTACAGACCTATGAGGAGTATG 3539  
 QY 3421 CATGCACTCTAGTCTTCTGATGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAG 3480  
 Db 3540 CATGCACTCTAGTCTTCTGATGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAG 3599  
 QY 3481 GAAGATCTAGTCTTCTGATGAGAAATGACATTAAGGAAAGTCTGCTGTTTACGAAAGC 3540  
 Db 3600 GAAGATCTAGTCTTCTGATGAGAAATGACATTAAGGAAAGTCTGCTGTTTACGAAAGC 3659  
 QY 3541 GTCAGAAAGAGAGCTTTACAGAGAGTCTAGCCCTTTACCCATACATATTTGGCTCAG 3600  
 Db 3660 GTCAGAAAGAGAGCTTTACAGAGAGTCTAGCCCTTTACCCATACATATTTGGCTCAG 3719  
 QY 3601 GGTACCGAAGAGAGGCGCAAGAAATTAAGTCTCTCAGAGAACTTATGAGTATGAGAT 3660  
 Db 3720 GGTACCGAAGAGAGGCGCGCAAGAAATTAAGTCTCTCAGAGAACTTATGAGTATGAGAT 3779  
 QY 3661 GAAGAGCTTCCCTGCTTCCAAACCTGTTATTTGGTAAAGTAAACAATATACCTTCAG 3720  
 Db 3780 GAAGAGCTTCCCTGCTTCCAAACCTGTTATTTGGTAAAGTAAACAATATACCTTCAG 3839  
 QY 3721 TCTACTAGGATAGACACCTGTTGCTACCGAGTCTGCTCTAAGAAACAGAGAGAAATTA 3780  
 Db 3840 TCTACTAGGATAGACACCTGTTGCTACCGAGTCTGCTCTAAGAAACAGAGAGAAATTA 3899  
 QY 3781 TTTATCATGGAAGATTAAGTCTTAATGACTGATACCAAGTATATTTGGCAAGGATCT 3840  
 Db 3900 TTTATCATGGAAGATTAAGTCTTAATGACTGATACCAAGTATATTTGGCAAGGATCT 3959  
 QY 3841 CAGGACATCACTTATGAGGAAACAAATGTTGCTAGTCTGTTTCTTACAGTGC 3900  
 Db 3960 CAGGACATCACTTATGAGGAAACAAATGTTGCTAGTCTGTTTCTTACAGTGC 4019  
 QY 3901 AGTAAATGGAAGTCTGATGCAAAATACAAACACCCAGAGTCTTCTGATTTGTTCT 3960  
 Db 4020 AGTAAATGGAAGTCTGATGCAAAATACAAACACCCAGAGTCTTCTGATTTGTTCT 4079  
 QY 3961 TCCAAACAAATGAGCATGCTGATGAAAGCCAGGAGTGTGCTGAGTCAAGAAATTG 4020  
 Db 4080 TCCAAACAAATGAGCATGCTGATGAAAGCCAGGAGTGTGCTGAGTCAAGAAATTG 4139

QY 4021 GTTTCAGATGATGAGAAAGAGAAACCGGCTTGGAAAGAAATATATCAAGAGACCAAGC 4080  
 Db 4140 GTTTCAGATGATGAGAAAGAGAAACCGGCTTGGAAAGAAATATATCAAGAGACCAAGC 4199  
 QY 4081 ATGGATTCAAACTTATGAGTGAAGCAGCATCTGGGTGTGAGAGTGAACAGGCTCTGAA 4140  
 Db 4200 ATGGATTCAAACTTATGAGTGAAGCAGCATCTGGGTGTGAGAGTGAACAGGCTCTGAA 4259  
 QY 4141 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAAACCTCAGCAGAGAGGATACATG 4200  
 Db 4260 GACTGCTCAGGGCTATCTCTCAGAGTGAATTTTAAACCTCAGCAGAGAGGATACATG 4319  
 QY 4201 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGATAGAGCTGTTTGAAGCAG 4260  
 Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGATAGAGCTGTTTGAAGCAG 4379  
 QY 4261 CATGGAGCCAGCCTTCTAAGTACAGTACCTTCCATCATATGATGACTCTTGCCTTGAG 4320  
 Db 4380 CATGGAGCCAGCCTTCTAAGTACAGTACCTTCCATCATATGATGACTCTTGCCTTGAG 4439  
 QY 4321 GACCTGGAAATCCAGAACAAAGCAGATCCAGAAAGCCTTCTGTCAGAGTTGAGTGTCT 4380  
 Db 4440 GACCTGGAAATCCAGAACAAAGCAGATCCAGAAAGCCTTCTGTCAGAGTTGAGTGTCT 4499  
 QY 4381 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGTCAGAGTTGAGTGTCT 4440  
 Db 4500 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGTCAGAGTTGAGTGTCT 4559  
 QY 4441 GCAGATAGTTCTACCACTTAATAAATAAGAACACAGAGTGAAGAGTATCCCTCTTAAA 4500  
 Db 4560 GCAGATAGTTCTACCACTTAATAAATAAGAACACAGAGTGAAGAGTATCCCTCTTAAA 4619  
 QY 4501 TGCCCATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
 Db 4620 TGCCCATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4679  
 QY 4561 AACTACCATCTCAAGAGAGAGCTCATTAAGTGTGATGATGATGATGATGATGATGATGAT 4620  
 Db 4680 AACTACCATCTCAAGAGAGAGCTCATTAAGTGTGATGATGATGATGATGATGATGATGAT 4739  
 QY 4621 GAGTCTGGGCGACAGCATTTTACGGAAGATCTTACTTGGCAAGGCAAGATCTAGAGGA 4680  
 Db 4740 GAGTCTGGGCGACAGCATTTTACGGAAGATCTTACTTGGCAAGGCAAGATCTAGAGGA 4799  
 QY 4681 ACCGCTTACCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 4740  
 Db 4800 ACCGCTTACCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 4859  
 QY 4741 GAAGACAGAGCCCGAGAGTACGTCGTTGGCAACATACATCTTCAACCTGTCATG 4800  
 Db 4860 GAAGACAGAGCCCGAGAGTACGTCGTTGGCAACATACATCTTCAACCTGTCATG 4919  
 QY 4801 AAGATTCGCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTGCTATCTACT 4860  
 Db 4920 AAGATTCGCCAATTTGAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTGCTATCTACT 4979  
 QY 4861 GATACGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAAAGCCAGAAATGACA 4920  
 Db 4980 GATACGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAAAGCCAGAAATGACA 5039  
 QY 4921 GCTTCAACAGAAAGGCTCAACAAAGATGCTCATGATGATGATGATGATGATGATGATGAT 4980  
 Db 5040 GCTTCAACAGAAAGGCTCAACAAAGATGCTCATGATGATGATGATGATGATGATGATGAT 5099  
 QY 4981 GAATTTATGCTCTGTATCAAGATTTGCCAGAAACACACATCACTTAATCAATCTAAT 5040  
 Db 5100 GAATTTATGCTCTGTATCAAGATTTGCCAGAAACACACATCACTTAATCAATCTAAT 5159  
 QY 5041 ACTGAAGAGCTACTCATGTTGTTATGAAGAAAGATGCTGATGTTGTTGTTGTTGTTGTTGTT 5100  
 Db 5160 ACTGAAGAGCTACTCATGTTGTTATGAAGAAAGATGCTGATGTTGTTGTTGTTGTTGTTGTT 5219

QY	5101	CTGAAATTTTCTTAGCAATTGCGGAGAGAAAATGGGAGTACTATTCTTGGGTAC	5160
Db	5220	CTGAATATTTTCTTAGCAATTGCGGAGAGAAAATGGGAGTACTATTCTTGGGTAC	5279
QY	5161	CAGTCATTTAAGAAAGAAAATGCTGATGAGCATGTTTGAAGTCAGAGAGATGTG	5220
Db	5380	CAGTCATTTAAGAAAGAAAATGCTGATGAGCATGTTTGAAGTCAGAGAGATGTG	5339
QY	5221	GTCATATGGAAGAACACCAACAGTGTCCAAAGGAGACAGAGAAATCCACGAGAGATGC	5280
Db	5340	GTCATATGGAAGAACACCAACAGTGTCCAAAGGAGACAGAGAAATCCACGAGAGATGC	5399
QY	5281	TTCAGGGGGCTAGAAATCTGTGTCTATGAGGCCCTTCACCAACATGCCACAGATCAACTG	5340
Db	5400	TTCAGGGGGCTAGAAATCTGTGTCTATGAGGCCCTTCACCAACATGCCACAGATCAACTG	5459
QY	5341	GAATGGATGGGACAGCTGTGTGGTGCTTCGTGTGGTGAAGAGGCTTATCATTTACCCCTT	5400
Db	5460	GAATGGATGGGACAGCTGTGTGGTGCTTCGTGTGGTGAAGAGGCTTATCATTTACCCCTT	5519
QY	5401	GGCAGAGGTGTCCACCCCAATTTGTGTTGTGCAGCCAGATGCTGGAGAGAGACAATGGC	5460
Db	5520	GGCAGAGGTGTCCACCCCAATTTGTGTTGTGCAGCCAGATGCTGGAGAGAGACAATGGC	5579
QY	5461	TTTCATGCAATTTGGGACAGATGTGTGAGGACCTGTGGTACCCGAGATGGGTGTGGAC	5520
Db	5580	TTTCATGCAATTTGGGACAGATGTGTGAGGACCTGTGGTACCCGAGATGGGTGTGGAC	5639
QY	5521	AGTGTAGCACTCTACCAAGTGGCAGAGAGCTGACACCTACCTGATACCCACAGATCCCCAC	5580
Db	5640	AGTGTAGCACTCTACCAAGTGGCAGAGAGCTGACACCTACCTGATACCCACAGATCCCCAC	5699
QY	5581	AGGCACATAC 5589	
Db	5700	AGGCACATAC 5708	

	RESULT	9	
AAV46460	ID	AAV46460 standard; cDNA; 5711 bp.	
XX	AC	AAV46460;	
XX	DT	18-NOV-1998 (first entry)	
XX	DE	Human BRCA1 omi2 polymorphism #3 cDNA.	
XX	KW	BRCA1, omi2; human; breast and ovarian cancer predisposing gene;	
KW	KW	polymorphism; susceptibility; anti-oncogene; tumour suppressor;	
KW	KW	chromosome 17q; ss.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	FT	CDS	120..5711
FT	FT	/**tag= a	/product= "BRCA1 omi2 protein"
FT	FT	variation	2731
FT	FT	/**tag= b	
FT	FT	/note= "this polymorphic variation can be a C or T nucleotide"	
XX	PN	US5750400-A.	
XX	PD	12-MAY-1998.	
XX	PF	12-FEB-1997;	97US-0798691.
XX	PR	12-FEB-1996;	96US-0598591.
XX	PA	12-FEB-1997;	97US-0798691.
XX		(ONCO-) INCORMED INC.	

PI	Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ:
PI	Scheller DB, Zeng B;
XX	
DR	WPI: 1998-296774/26.
XX	
PT	BRCA1 om1 gene coding sequences - useful for distinguishing between
PT	polymorphisms and mutation(s) in the screening for disposition to
PT	breast or ovarian cancer
XX	
PS	Claim 2e: Page -: 54pp; English.
XX	
CC	This sequence encodes a human BRCA1 (breast and ovarian cancer
CC	predisposing gene) om2 gene in which a polymorphic variation occurs at
CC	nucleotide 2733. This sequence and other polymorphic variations of this
CC	sequence are useful for the identification of an individual who may or
CC	may not have an increased susceptibility to breast or ovarian cancer.
CC	The sequences used identify gene changes which are due to polymorphisms,
CC	rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC	suppressor) which is involved in genetic inheritance of cancers,
CC	especially breast and ovarian cancer. It is found at human chromosome
CC	17q which is known to be linked to cancer susceptibility, especially
CC	breast cancer. Cells containing a mutation in this gene lose the
CC	wild-type function of BRCA1 and are more susceptible to cancers.
CC	NOTE: This sequence does not appear in the specification but has been
CC	created from the wild type BRCA1 om2 gene represented in AAV6449.
XX	
XX	
XX	
Sequence	5711 BP; 1956 A; 1098 C; 1274 G; 1382 T; 1 other;

Query Match	100.0%;	Score 5587;	DB 19;	Length 5711;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 5587; Conservative	1;	Mismatches	1;	Indels 0; Gaps 0;

QY	1	ATGATATTATTCGCTTCCTTGGCTTGAAGATACAAAGATGATCATTAATGCTATGACAGAA	60
Db	120	ATGATATTATTCGCTTCCTTGGCTTGAAGATACAAAGATGATCATTAATGCTATGACAGAA	179
QY	61	ATCTAGAGTGTCCATCTGTCTGGAGTTGATCAAGAACCTGTCTCCACAAAGTGTAC	120
Db	180	ATCTAGAGTGTCCATCTGTCTGGAGTTGATCAAGAACCTGTCTCCACAAAGTGTAC	239
QY	121	CACATATTTTGGCAATTTTGGCATGTGTAACCTTCCAACCGAAGAAAGGCGCTTCACAG	180
Db	240	CACATATTTTGGCAATTTTGGCATGTGTAACCTTCCAACCGAAGAAAGGCGCTTCACAG	299
QY	181	TGTCCCTTATGTAAAGATGATATTAACCAAAAGAGCGCTACAGAAATAGAGATTAGT	240
Db	300	TGTCCCTTATGTAAAGATGATATTAACCAAAAGAGCGCTACAGAAATAGAGATTAGT	359
QY	241	CAACTGTGTGAAGAGCTATTTGAAATVCATTTGTGCTTTACGTTGACACAGGTTTGGAG	300
Db	360	CAACTGTGTGAAGAGCTATTTGAAATVCATTTGTGCTTTACGTTGACACAGGTTTGGAG	419
QY	301	TATGCAACAGCTATTAATTTTGCACAAAAGGAAATACTCTCTGAACTCTAAAGAAAT	360
Db	420	TATGCAACAGCTATTAATTTTGCACAAAAGGAAATACTCTCTGAACTCTAAAGAAAT	479
QY	361	GAAGTTTCTATCATCTCCAAAGTATGGGCTACACAAACCGTGCACAAACATCTTACAGAGT	420
Db	480	GAAGTTTCTATCATCTCCAAAGTATGGGCTACACAAACCGTGCACAAACATCTTACAGAGT	539
QY	421	GAAACCCGAAATTCCTCTTGCAGGAACAACAGTCTCAGTGTCCAACTCTTAACTTTGA	480
Db	540	GAAACCCGAAATTCCTCTTGCAGGAACAACAGTCTCAGTGTCCAACTCTTAACTTTGA	599
QY	481	ACTGTGAGAACTCTGAGGACAAAGCAGGGGATACAACCTCAAAAAGACGTGTCTACATT	540
Db	600	ACTGTGAGAACTCTGAGGACAAAGCAGGGGATACAACCTCAAAAAGACGTGTCTACATT	659
QY	541	GAATTTGGATCTGATCTCTTGGAAACATTCACCTTAATTAAGCAACTTTTGCATGTGGGA	600
Db	660	GAATTTGGATCTGATCTCTTGGAAACATTCACCTTAATTAAGCAACTTTTGCATGTGGGA	719
QY	601	GATCAAGAAATTTTCAAAATCACCCCTCAGAGAACCAAGGATGAATCACTTTGGATTC	660

|||||  
Db 720 GATCAGAAATGTTACAAATCACCCCTCAAGAACAGGATGAATTCAGTTGGATTCT 779  
QY 661 GCAAAAAAGGCTGCTGTGAAATTTTCTGAGACGATGTAACAAATCTGAACATCATCA 720  
Db 780 GCAAAAAAGGCTGCTGTGAAATTTTCTGAGACGATGTAACAAATCTGAACATCATCA 839  
QY 721 CCCAGTAATTAATGATTGAACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 840 CCCAGTAATTAATGATTGAACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 899  
QY 781 TATCAGGATGATTCTGTTTCAAACTTGATGTGAGGACATGGCACAATTAATCTGATGCC 840  
Db 900 TATCAGGATGATTCTGTTTCAAACTTGATGTGAGGACATGGCACAATTAATCTGATGCC 959  
QY 841 AGCTCATATACAGATGAGAACACACAGTTTATTAATCTCACTAAGACAGATGATGAA 900  
Db 960 AGCTCATATACAGATGAGAACACACAGTTTATTAATCTCACTAAGACAGATGATGAA 1019  
QY 901 AAGGCTGAATTTCTGTATATAAAGCAACACCTGGCTTACAGAGAGCCAACTAATCAGA 960  
Db 1020 AAGGCTGAATTTCTGTATATAAAGCAACACCTGGCTTACAGAGAGCCAACTAATCAGA 1079  
QY 961 TGGGCTGGAAGTAGAAGAAACATGTAATGATAGCGGACCTCCAGACAGAAAAAGGTA 1020  
Db 1080 TGGGCTGGAAGTAGAAGAAACATGTAATGATAGCGGACCTCCAGACAGAAAAAGGTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAAGATGAATAGCAGAACTGCCATGC 1080  
Db 1140 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAAGATGAATAGCAGAACTGCCATGC 1199  
QY 1081 TCAGAGAAATCTGTAGATGATGAGATGTTCTGTGATTAACATAATATAGCAGATTCAG 1140  
Db 1200 TCAGAGAAATCTGTAGATGATGAGATGTTCTGTGATTAACATAATATAGCAGATTCAG 1259  
QY 1141 AAGGTAATGATGTTTTCAGAACTGATGAACCTGTTAGCTTGATGATCAGATGAT 1200  
Db 1260 AAGGTAATGATGTTTTCAGAACTGATGAACCTGTTAGCTTGATGATCAGATGAT 1319  
QY 1201 GGGGATCTGTAATCAATGCAAGTAAAGTATGATGATGATGATGATGATGATGATGAT 1260  
Db 1320 GGGGATCTGTAATCAATGCAAGTAAAGTATGATGATGATGATGATGATGATGATGAT 1379  
QY 1261 GAATATTTCTGTTCTTCAAGAGAAATAGACTTACTGCGCAGTATCTCTGAGGCTTGA 1320  
Db 1380 GAATATTTCTGTTCTTCAAGAGAAATAGACTTACTGCGCAGTATCTCTGAGGCTTGA 1439  
QY 1321 ATATGTAAGAGTAAAGAGTTCATCCAAATCAGTAGAGATTAATTTGAAGCAAAATA 1380  
Db 1440 ATATGTAAGAGTAAAGAGTTCATCCAAATCAGTAGAGATTAATTTGAAGCAAAATA 1499  
QY 1381 TTTGGAAAAACCTATGGAAGAGGCAAGGCTCCCACTTAAGCCATGATGATAAAT 1440  
Db 1500 TTTGGAAAAACCTATGGAAGAGGCAAGGCTCCCACTTAAGCCATGATGATAAAT 1559  
QY 1441 CTATATATAGAGCATTTTCTTACTGAGCCACAGATATATACAAAGGCTCCCTCACAAT 1500  
Db 1560 CTATATATAGAGCATTTTCTTACTGAGCCACAGATATATACAAAGGCTCCCTCACAAT 1619  
QY 1501 AAATTAAGCGTAAAGAGAGACCTACATGAGGCTTCATCTGAGATTTTATCAAGAAA 1560  
Db 1620 AAATTAAGCGTAAAGAGAGACCTACATGAGGCTTCATCTGAGATTTTATCAAGAAA 1679  
QY 1561 GCGATTTGGAGTTCAAAAAGACTCCTGAAATATATAATCAGGAGACTAACCAAGGAG 1620  
Db 1680 GCGATTTGGAGTTCAAAAAGACTCCTGAAATATATAATCAGGAGACTAACCAAGGAG 1739  
QY 1621 CAGAAATGCTCAAGTGAATATTTACTAATAGTGTGATGAGAAATTAACCAAAAGGTGAT 1680  
Db 1740 CAGAAATGCTCAAGTGAATATTTACTAATAGTGTGATGAGAAATTAACCAAAAGGTGAT 1799  
QY 1681 TCTATTGAGATGAGAAAAATCTTAACCAATAGAAATCACTGAAAAAGAAATCTGTTTC 1740  
|||||

Db 1800 TCTATTGAGATGAGAAAAATCTTAACCAATAGAAATCACTGAAAAAGAAATCTGTTTC 1859  
QY 1741 AAAACGAAACCTGACCTATATAGACAGATATAAGCAATATGAACTGCAATTAATATC 1800  
Db 1860 AAAACGAAACCTGACCTATATAGACAGATATAAGCAATATGAACTGCAATTAATATC 1919  
QY 1801 CACAATTCAAAGACACCTTAATAAAGAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
Db 1920 CACAATTCAAAGACACCTTAATAAAGAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1979  
QY 1861 CATGCGCTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920  
Db 1980 CATGCGCTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2039  
QY 1921 ATTGATAGTTGTTCTAGCATGAAAGATTAAGAAAAAGTAAACCAATATGCGCATG 1980  
Db 2040 ATTGATAGTTGTTCTAGCATGAAAGATTAAGAAAAAGTAAACCAATATGCGCATG 2099  
QY 1981 AGGACAGCAGAAACCTCAACTCATGAGAGTAAAGAACTGCAACTGAGGAGCAAGAG 2040  
Db 2100 AGGACAGCAGAAACCTCAACTCATGAGAGTAAAGAACTGCAACTGAGGAGCAAGAG 2159  
QY 2041 AGTAACAGCCAAATGAACAGACAGTAAAGACATGACAGATGATCTTTCCAGAGCTG 2100  
Db 2160 AGTAACAGCCAAATGAACAGACAGTAAAGACATGACAGATGATCTTTCCAGAGCTG 2219  
QY 2101 AAGTTAAACAATGACCTGCTGTTTCTTACTAATGATGTTCAATATACATGATTAAGAA 2160  
Db 2220 AAGTTAAACAATGACCTGCTGTTTCTTACTAATGATGTTCAATATACATGATTAAGAA 2279  
QY 2161 TTTGTCATCTTACCTTCCAGAGAAAGAAAGAAAGAAAGTAAAGTAAAGT 2220  
Db 2280 TTTGTCATCTTACCTTCCAGAGAAAGAAAGAAAGAAAGTAAAGTAAAGT 2339  
QY 2221 TCTAATATCTGTAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTGCACACT 2280  
Db 2340 TCTAATATCTGTAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTGCACACT 2399  
QY 2281 GAAAGATCTGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
Db 2400 GAAAGATCTGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2459  
QY 2341 GAAAGATCTGTGTACTAGAGAGTATGACACTCTAGGAGAGCAAAACAGAACCAATATA 2400  
Db 2460 GAAAGATCTGTGTACTAGAGAGTATGACACTCTAGGAGAGCAAAACAGAACCAATATA 2519  
QY 2401 TGTGTAGTCAAGTGTGACAGATTTGAAGAACCCCAAGGACTAATTCATGTTGTTCCAA 2460  
Db 2520 TGTGTAGTCAAGTGTGACAGATTTGAAGAACCCCAAGGACTAATTCATGTTGTTCCAA 2579  
QY 2461 GATTAATAGAAATGACACAGAGCTTAAATATCCATTTGGGACATGAAAGTTAACCCACT 2520  
Db 2580 GATTAATAGAAATGACACAGAGCTTAAATATCCATTTGGGACATGAAAGTTAACCCACT 2639  
QY 2521 CCGGAAAAAGCAATAGAAATGGAAGAAAGTGAAGTATGATGATGATGATGATGATGATG 2580  
Db 2640 CCGGAAAAAGCAATAGAAATGGAAGAAAGTGAAGTATGATGATGATGATGATGATGATG 2699  
QY 2581 TTCAAGTTTCAAGGCGCAGTATTTGCTCTGTTTCAATCCAGAAATGAGAAAGAG 2640  
Db 2700 TTCAAGTTTCAAGGCGCAGTATTTGCTCTGTTTCAATCCAGAAATGAGAAAGAG 2759  
QY 2641 GAATGTGCAACATCTCTGCGCACCTGCGGCTTTAAAGAAACCAAAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGCAACATCTCTGCGCACCTGCGGCTTTAAAGAAACCAAAAGTCCAAAGTCACT 2819  
QY 2701 TTTGAATGTGAACAAAGAGAAATCAAGAGAAAGTAAAGTAAATATCAAGCCCTGTA 2760  
Db 2820 TTTGAATGTGAACAAAGAGAAATCAAGAGAAAGTAAAGTAAATATCAAGCCCTGTA 2879  
QY 2761 CAGACAGTTAATATCACTGAGGCTTTCTGTTGTTGTCAGAAAGTAAAGCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCACTGAGGCTTTCTGTTGTTGTCAGAAAGTAAAGCAGTTGAT 2939  
|||||

QY	2821	AATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTCTAGAGGC	2880
Db	2940	AATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTCTAGAGGC	2999
QY	2881	AACGAAACTGGACCTATTACTCCAAATTAACATGACCTTTTACAAACCCATATCGTATA	2940
Db	3000	AACGAAACTGGACCTATTACTCCAAATTAACATGACCTTTTACAAACCCATATCGTATA	3059
QY	2941	CCACACATTTTCCCATCAAGTCATTTGTAAAACATAATGTAAAGAAAATCTGCTAGAG	3000
Db	3060	CCACACATTTTCCCATCAAGTCATTTGTAAAACATAATGTAAAGAAAATCTGCTAGAG	3119
QY	3001	GAAAACTTTGAGGAAACATTCAATGTGCACCTGAAAGAGAAATGGGAAATGAGAACATTCCA	3060
Db	3120	GAAAACTTTGAGGAAACATTCAATGTGCACCTGAAAGAGAAATGGGAAATGAGAACATTCCA	3179
QY	3061	AGTACAGTGAGCACAAATTAGCCGTATTAACATTAGAGAAAATGTTTTTAAAGAACCCAGC	3120
Db	3180	AGTACAGTGAGCACAAATTAGCCGTATTAACATTAGAGAAAATGTTTTTAAAGAACCCAGC	3239
QY	3121	TCACCAATATTATTAAGAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTATGAA	3180
Db	3240	TCACCAATATTATTAAGAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTATGAA	3299
QY	3181	ATAGTTCACAGTATGAAGAAACATTCACACAGACAGTATGAACACAGGGGCCAAAATNTG	3240
Db	3300	ATAGTTCACAGTATGAAGAAACATTCACACAGACAGTATGAACACAGGGGCCAAAATNTG	3359
QY	3241	AATGCTATGCTTAGTATGATGAGGGGTTTGGACCTGAGGCTATTAACAAGCTCTTCGTGA	3300
Db	3360	AATGCTATGCTTAGTATGATGAGGGGTTTGGACCTGAGGCTATTAACAAGCTCTTCGTGA	3419
QY	3301	AGTAAATTGTAAGCATCTCGAAATTAAGAAACAGACATATGAAGAGTAGTTCAGACTGTT	3360
Db	3420	AGTAAATTGTAAGCATCTCGAAATTAAGAAACAGACATATGAAGAGTAGTTCAGACTGTT	3479
QY	3361	AATACAGATTTCTCTCCATATCTGATTTCAATTAACATTAAGAACAGCCATGGGAAAGTAGT	3420
Db	3480	AATACAGATTTCTCTCCATATCTGATTTCAATTAACATTAAGAACAGCCATGGGAAAGTAGT	3539
QY	3421	CATGCATCTCAGGTTTGTCTTGAGACACCTGATGACCTGTAGATGATGGTAATTAAG	3480
Db	3540	CATGCATCTCAGGTTTGTCTTGAGACACCTGATGACCTGTAGATGATGGTAATTAAG	3599
QY	3481	GAAGTACTACTGTTTGGTGAAGAAATGACATTAAAGAACTTCTGCTGTTTTAGCAAAAGC	3540
Db	3600	GAAGTACTACTGTTTGGTGAAGAAATGACATTAAAGAACTTCTGCTGTTTTAGCAAAAGC	3659
QY	3541	GTCCAGAAAGAGAGACTTAGCAGAGAGTCTTAGCCCTTTCACCCATPACACATTTGGCTCAG	3600
Db	3660	GTCCAGAAAGAGAGACTTAGCAGAGAGTCTTAGCCCTTTCACCCATPACACATTTGGCTCAG	3719
QY	3601	GTTTACCGAAGAGGGGCCAAGAATTTAGAGTCTCTAGAGAGAACTTATCTAGTAGAGAT	3660
Db	3720	GTTTACCGAAGAGGGGCCAAGAATTTAGAGTCTCTAGAGAGAACTTATCTAGTAGAGAT	3779
QY	3661	GAAGAGCTTCCCTGCTTCCAACTTGTATTGTTGGTAAAGTAAACAATATACCTTCTCAG	3720
Db	3780	GAAGAGCTTCCCTGCTTCCAACTTGTATTGTTGGTAAAGTAAACAATATACCTTCTCAG	3839
QY	3721	TCTACTAGGCAATACACCGTTGCTACCGAGTGTCTGCTTAAGACACAGAGGAGAAATTTA	3780
Db	3840	TCTACTAGGCAATACACCGTTGCTACCGAGTGTCTGCTTAAGACACAGAGGAGAAATTTA	3899
QY	3781	TTATCATTTGAAGAAATAGCTTAAATGACTGTACGTAAACACAGGTAATATTGGCAAGGCATCT	3840
Db	3900	TTATCATTTGAAGAAATAGCTTAAATGACTGTACGTAAACACAGGTAATATTGGCAAGGCATCT	3959
QY	3841	CAGGAACATCACCTTAGTACGAGAAACAATAATGTTCTGCTAGCTGTTTTCTTCACAGTGC	3900
Db	3960	CAGGAACATCACCTTAGTACGAGAAACAATAATGTTCTGCTAGCTGTTTTCTTCACAGTGC	4019

OY	3901	AGTGAATTGGAAAGACTTGACTGCTGCAATACAAACACCCAGGATCCCTTCTTGATGTGGTCT	3960
Db	4020	AGTGAATTGGAAAGACTTGACTGCTGCAATACAAACACCCAGGATCCCTTCTTGATGTGGTCT	4079
OY	3961	TCCAACAATAATGAGCATCAGTCTGTAAGGCGAGGAGTTGGTCTGAGTACAGAAGATTG	4020
Db	4080	TCCAACAATAATGAGCATCAGTCTGTAAGGCGAGGAGTTGGTCTGAGTACAGAAGATTG	4139
OY	4021	GTTTTCAGATGATGAAGAAAGAGAAACGGGCTTGGAAGAAATATCAAGAAAGCAAGC	4080
Db	4140	GTTTTCAGATGATGAAGAAAGAGAAACGGGCTTGGAAGAAATATCAAGAAAGCAAGC	4199
OY	4081	ATGGATTCAAACTTATAGTGTGAGAGCACCATCTGGGTGTGAGAGTGAAGAACAGCTCTCTAA	4140
Db	4200	ATGGATTCAAACTTATAGTGTGAGAGCACCATCTGGGTGTGAGAGTGAAGAACAGCTCTCTAA	4259
OY	4141	GACTCTCAGGGCTATCCCTCTCAGAGTACATTTTAACCACTCAGCAAGGATACCATG	4200
Db	4260	GACTCTCAGGGCTATCCCTCTCAGAGTACATTTTAACCACTCAGCAAGGATACCATG	4319
OY	4201	CAACATTAACCTGATTAAGCTCCAGCAGGAAATAGCTGATAGAAAGCTGTGTAGAACAG	4260
Db	4320	CAACATTAACCTGATTAAGCTCCAGCAGGAAATAGCTGATAGAAAGCTGTGTAGAACAG	4379
OY	4261	CATGGGAGCCAGCCCTTCTTAACAGCTACCCCTTCATATAAGTACCTTCTTGCCCTTAG	4320
Db	4380	CATGGGAGCCAGCCCTTCTTAACAGCTACCCCTTCATATAAGTACCTTCTTGCCCTTAG	4439
OY	4321	GACCTGCGAAATCCCGAACAAGCAGCATCAGAAAGCAGTATTAATTAATTCACAGAAAGT	4380
Db	4440	GACCTGCGAAATCCCGAACAAGCAGCATCAGAAAGCAGTATTAATTAATTCACAGAAAGT	4499
OY	4381	AGTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAATTTGAGTGTCT	4440
Db	4500	AGTGAATACCCCTATTAAGCCAGAAATCCAGAAAGGCTTCTGCTGACAAATTTGAGTGTCT	4559
OY	4441	GCAGTACTTCTACCAAGTAAAAATTAACAACAGAGAGTGAAGGTCATCCCTCTTAA	4500
Db	4560	GCAGTACTTCTACCAAGTAAAAATTAACAACAGAGAGTGAAGGTCATCCCTCTTAA	4619
OY	4501	TGCCCATATTATGATAGTAGGAGTACATGCAACATTTGCTGGGAGCTTCAGAAATGA	4560
Db	4620	TGCCCATATTATGATAGTAGGAGTACATGCAACATTTGCTGGGAGCTTCAGAAATGA	4679
OY	4561	AACTAACCATCTCAAGAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGGAA	4620
Db	4680	AACTAACCATCTCAAGAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGGAA	4739
OY	4621	GAGTGTGGGCGACACGATTTTGACGAAACATCTTACCTGCCAAGGCAAGATCTAGAGGA	4680
Db	4740	GAGTGTGGGCGACACGATTTTGACGAAACATCTTACCTGCCAAGGCAAGATCTAGAGGA	4799
OY	4681	ACCCTTACCTGGAACTCTGSAATCGGCTCTCTGATGAGCCGGAATCTGATCCCTCT	4740
Db	4800	ACCCTTACCTGGAACTCTGSAATCGGCTCTCTGATGAGCCGGAATCTGATCCCTCT	4859
OY	4741	GAAGACAGAGCCCCAGAGTCAAGCTCTGTTGGCAACATACCATCTTCAACCTTCGATTG	4800
Db	4860	GAAGACAGAGCCCCAGAGTCAAGCTCTGTTGGCAACATACCATCTTCAACCTTCGATTG	4919
OY	4801	AAAGTTCCTCCCAATTTGAAGTTTGACAGAACTCTCCCAAGAGTCCAGCTGCTCATATACT	4860
Db	4920	AAAGTTCCTCCCAATTTGAAGTTTGACAGAACTCTCCCAAGAGTCCAGCTGCTCATATACT	4979
OY	4861	GATACCTGTGGGTATTAATGCAATGGAGAAAGTGTGAGCAGGGAGAAAGCCAGATTGACA	4920
Db	4980	GATACCTGTGGGTATTAATGCAATGGAGAAAGTGTGAGCAGGGAGAAAGCCAGATTGACA	5039
OY	4921	GCTTCAACAGAAAGGCTCAACAAAAGATGTCCATGAGTGTGGCTGTGACCCCGAA	4980
Db	5040	GCTTCAACAGAAAGGCTCAACAAAAGATGTCCATGAGTGTGGCTGTGACCCCGAA	5099
OY	4981	GAATTTATGCTCGTGTACAACTTTTGCCGAAGAACACCATCACTTAATCAATCTAAT	5040

Db	5100	GAATTTATGCGCGGTACAAAGTTTGGCCAGAAACACCACATCACTTAACTAAATCTAATT	5159
QY	5041	ACTGAAGAGCTACTCATGTTGTTATGAAAAACAGATCTGAGTTTGTGTGAACGACA	5100
Db	5160	ACTGAAGAGCTACTCATGTTGTTATGAAAAACAGATCTGAGTTTGTGTGAACGACA	5219
QY	5101	CTGAAATATTTTTCAGGAATGCGGGAGGAAAAATGGGAGTACTATTTCTGGGTACC	5160
Db	5220	CTGAAATATTTTTCAGGAATGCGGGAGGAAAAATGGGAGTACTATTTCTGGGTACC	5279
QY	5161	CAGTCTATTTAAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG	5220
Db	5280	CAGTCTATTTAAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG	5339
QY	5221	GTCATATGGAAGAAACCCACCAAGGTCCAAAGCGACAGAGAGATATCCAGAGACAGAAAGATC	5280
Db	5340	GTCATATGGAAGAAACCCACCAAGGTCCAAAGCGACAGAGAGATATCCAGAGACAGAAAGATC	5399
QY	5281	TTTCAGGGGGGTAGAAATCTGTGCTATGAGGCGCTTCACCAACATGCCACAGATCACTG	5340
Db	5400	TTTCAGGGGGGTAGAAATCTGTGCTATGAGGCGCTTCACCAACATGCCACAGATCACTG	5459
QY	5341	GAATGATGGGTACAGCTGTGTGTGCTTCTGTGTGTGAAGAGCTTTCATCATTCACCCCTT	5400
Db	5460	GAATGATGGGTACAGCTGTGTGTGCTTCTGTGTGTGAAGAGCTTTCATCATTCACCCCTT	5519
QY	5401	GGCAGAGGTGCCACCCCAATTTGTGTGTGTGAGCAGCATGCTGTGAGACAGAGGACAATGGC	5460
Db	5520	GGCAGAGGTGCCACCCCAATTTGTGTGTGTGAGCAGCATGCTGTGAGACAGAGGACAATGGC	5579
QY	5461	TTCCATGCAATTGGGGCAGATGTGTGAGGACACCTGTGTGTGACCCGAGAGTGGGTGTTGGAC	5520
Db	5580	TTCCATGCAATTGGGGCAGATGTGTGAGGACACCTGTGTGTGACCCGAGAGTGGGTGTTGGAC	5639
QY	5521	AGGTATACACACTCAACGAGTGGCAGGAGCTGGAGACCTGATGATACCCAGATCCCCCACC	5580
Db	5640	AGGTATACACACTCAACGAGTGGCAGGAGCTGGAGACCTGATGATACCCAGATCCCCCACC	5699
QY	5581	AGGCACACTAC 5589	
Db	5700	AGGCACACTAC 5708	
RESULT 10			
AAV46461			
ID	AAV46461	standard; cDNA; 5711 BP.	
XX	AAV46461;		
AC	18-NOV-1998	(first entry)	
XX			
DT			
XX			
DE	Human BRCA1 omi2 polymorphism #4 cDNA.		
XX			
KM	BRCA1; omi2; human; breast and ovarian cancer predisposing gene;		
KW	polymorphism; susceptibility; anti-oncogene; tumour suppressor;		
KW	chromosome 17q; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	location/Qualifiers	
FT	CDS	120..5711	
FT		/*tag= a	
FT		/product= "BRCA1 omi2 protein"	
FT	variation	3232	
FT		/*tag= b	
FT		/note= "this polymorphic variation can be an A or G	
XX		nucleotide"	
XX	US5750400-A.		
XX	12-MAY-1998.		
XX			

XX 12-FEB-1997; 97US-0798691.  
PR 12-FEB-1996; 96US-0598591.  
PR 12-FEB-1997; 97US-0798691.  
XX  
XX (ONCO-) ONCORMED INC.  
XX  
XX  
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;  
PI Schelter DB, Zeng B;  
XX  
XX WPI; 1998-296774/26.  
DR  
PT BRCA1 omi gene coding sequences - useful for distinguishing between  
PT polymorphisms and mutation(s) in the screening for disposition to  
PT breast or ovarian cancer  
XX  
PS  
XX  
XX Claim 2e; Page -: 54pp; English.

This sequence encodes a human BRCA1 (breast and ovarian cancer predisposing gene) omi2 gene in which a polymorphic variation occurs at nucleotide 3332. This sequence and other polymorphic variations of this sequence are useful for the identification of an individual who may or may not have an increased susceptibility to breast or ovarian cancer. The sequences used identify gene changes which are due to polymorphisms, rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour suppressor) which is involved in genetic inheritance of cancers, especially breast and ovarian cancer. It is found at human chromosome 17q which is known to be linked to cancer susceptibility, especially breast cancer. Cells containing a mutation in this gene lose the wild-type function of BRCA1 and are more susceptible to cancers. NOTE: This sequence does not appear in the specification but has been created from the wild type BRCA1 omi2 gene represented in AA046649.

Sequence 5711 BP; 1995 A; 1098 C; 1274 G; 1383 T; 1 other;

Query Match	100.0%	Score 5587;	DB 19;	Length 5711;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 5587;	Conservative	1;	Mismatches	1; Indels 0; Gaps 0;
QY	1	ATGATATTTATTCGCTCTTCGCGTTTGAGAACGTACAAAATGTCAATTAATGCTATGACGAAA	60	
Db	120	ATGATATTTATTCGCTCTTCGCGTTTGAGAACGTACAAAATGTCAATTAATGCTATGACGAAA	179	
QY	61	ATCTTAGAGTCTCCCATCTGCTGTGGAGTGAATCAAGGAACCTGCTCCACAAAGTGTCAC	120	
Db	180	ATCTTAGAGTCTCCCATCTGCTGTGGAGTGAATCAAGGAACCTGCTCCACAAAGTGTCAC	239	
QY	121	CACATATTTTGCAAATTTTGCATGCTGTAACCTTCCCAACGAGAAAGAGGCGCTTCACAG	180	
Db	240	CACATATTTTGCAAATTTTGCATGCTGTAACCTTCCCAACGAGAAAGAGGCGCTTCACAG	299	
QY	181	TGTCCTTTATGTAAAGATGATATAACCAAAAGAGCGCTTACAGAAAGTACGAGATTTAGT	240	
Db	300	TGTCCTTTATGTAAAGATGATATAACCAAAAGAGCGCTTACAGAAAGTACGAGATTTAGT	359	
QY	241	CAACTGTGTGAAGAGCTATGTGAAAAATCATTTGTGCTTTCAGCTTGATACAGAGTTTGAG	300	
Db	360	CAACTGTGTGAAGAGCTATGTGAAAAATCATTTGTGCTTTCAGCTTGATACAGAGTTTGAG	419	
QY	301	TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAAAATAACTCTCTGGAACATCTAAAGAT	360	
Db	420	TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAAAATAACTCTCTGGAACATCTAAAGAT	479	
QY	361	GAACTTTCTATCATATCCAAAGTATGGGCTACAGAAAACGCTGGCAAAAGACTCTTACAGAGT	420	
Db	480	GAACTTTCTATCATATCCAAAGTATGGGCTACAGAAAACGCTGGCAAAAGACTCTTACAGAGT	539	
QY	421	GAAACCGGAAATCCTTCCCTTCGACAGAAACAGCTGCAATGTCCAACTCTCTAACCTTGGA	480	
Db	540	GAAACCGGAAATCCTTCCCTTCGACAGAAACAGCTGCAATGTCTCAACTCTCTAACCTTGGA	599	
QY	481	ACTGTGAACTCTGAGAGCAAAAGCAGCGGATACAACTCAAAAGAGCTGTCTTCACTT	540	



```
Db 600 ACTGTGAGACTCTGTGAGCAACAAAGCGGATACAACTCTCAAAAGACGTCTGTACATT 659
Qy 541 GAATGGGATCTGATCTTCTGAGATACCGTTAATAAGCAACTATTGACAGTGGGA 600
Db 660 GAATGGGATCTGATCTTCTGAGATACCGTTAATAAGCAACTATTGACAGTGGGA 719
Qy 601 GATCAAGAAATTTGTACAAATCACCCTCAAGAACGAGGATGAAATCAGTTTGATCT 660
Db 720 GATCAAGAAATTTGTACAAATCACCCTCAAGAACGAGGATGAAATCAGTTTGATCT 779
Qy 661 GAAAAAAGGCTGCTGTGAAATTTCTGAGAGGATGTAACAATCTGAACTCATCA 720
Db 780 GAAAAAAGGCTGCTGTGAAATTTCTGAGAGGATGTAACAATCTGAACTCATCA 839
Qy 721 CCCAGTAATATGATTTGGAACACCACTGAGAACGCTGACAGCTGAGAGCATCCAGAAAG 780
Db 840 CCCAGTAATATGATTTGGAACACCACTGAGAACGCTGACAGCTGAGAGCATCCAGAAAG 899
Qy 781 TATCAGGGTAGTTCCTGTTTCAAACTTGATGGAGCCATGTGGCAACAATCTCATGCC 840
Db 900 TATCAGGGTAGTTCCTGTTTCAAACTTGATGGAGCCATGTGGCAACAATCTCATGCC 959
Qy 841 ACCTATTACAGCATGAGAACGCACTTATTACTACTAATAAGACGAATGATGTAGAA 900
Db 960 ACCTATTACAGCATGAGAACGCACTTATTACTACTAATAAGACGAATGATGTAGAA 1019
Qy 901 AAGGCTGAATTCGTATATAAAGCAAAACAGCGCTGCTTAGCAAGAGCCAACTAATCAGA 960
Db 1020 AAGGCTGAATTCGTATATAAAGCAAAACAGCGCTGCTTAGCAAGAGCCAACTAATCAGA 1079
Qy 961 TGGGCTGGAAGTAAGGAACATGTATATAGGCGGACTCCGACACAGAAAAAAGTA 1020
Db 1080 TGGGCTGGAAGTAAGGAACATGTATATAGGCGGACTCCGACACAGAAAAAAGTA 1139
Qy 1021 GATCGAATGCTGATCCCTGTGTGAGAAAAAAGTAATGATAGAGAACTGCCATGC 1080
Db 1140 GATCGAATGCTGATCCCTGTGTGAGAAAAAAGTAATGATAGAGAACTGCCATGC 1199
Qy 1081 TCAGAGAATCCTAGAGATCTAGAGATGTTCTTGATTAACATAAATAGCAGCATTCAG 1140
Db 1200 TCAGAGAATCCTAGAGATCTAGAGATGTTCTTGATTAACATAAATAGCAGCATTCAG 1259
Qy 1141 AAAGTAAATGAGTGTTCCTGCAAGAGTGAATGAACTTTAGGTTCTATGACTACATGAT 1200
Db 1260 AAAGTAAATGAGTGTTCCTGCAAGAGTGAATGAACTTTAGGTTCTATGACTACATGAT 1319
Qy 1201 GGGGAGTCTGAATCAAAATCCCAAAGTAGTGAATGTTGAGAGCTTAAATGAGTAGAT 1260
Db 1320 GGGGAGTCTGAATCAAAATCCCAAAGTAGTGAATGTTGAGAGCTTAAATGAGTAGAT 1379
Qy 1261 GAATATTCGTGTTCTTCAGAGAAATAGACTTACTGGCCAGTAGATCCTCATGAGGCTTTA 1320
Db 1380 GAATATTCGTGTTCTTCAGAGAAATAGACTTACTGGCCAGTAGATCCTCATGAGGCTTTA 1439
Qy 1321 ATATGTAAAGTGAAGAGTTCACATCCAAATCAGTAGAGATATATTGAAGCAAAATA 1380
Db 1440 ATATGTAAAGTGAAGAGTTCACATCCAAATCAGTAGAGATATATTGAAGCAAAATA 1499
Qy 1381 TTTGGGAAACCTATCGGAGAAAGGCAAGCCTCCCAACTTAAGCATGTAATGTAAT 1440
Db 1500 TTTGGGAAACCTATCGGAGAAAGGCAAGCCTCCCAACTTAAGCATGTAATGTAAT 1559
Qy 1441 CTAATTTATAGAGACTTTGTTACTAGCCACAGATATATACAGAGGCTCCCTCACAAAT 1500
Db 1560 CTAATTTATAGAGACTTTGTTACTAGCCACAGATATATACAGAGGCTCCCTCACAAAT 1619
Qy 1501 AAATTTAAAGCTAAAGAGACCTACATCAGGCTTCATCTGAGAGATTTATCAAGAA 1560
Db 1620 AAATTTAAAGCTAAAGAGACCTACATCAGGCTTCATCTGAGAGATTTATCAAGAA 1679
Qy 1561 GCAGATTTGGAGCTCAAAAGACCTCTGAAATGATTAATCAGGAACTAACCAAGCGAG 1620
Db 1680 GCAGATTTGGAGCTCAAAAGACCTCTGAAATGATTAATCAGGAACTAACCAAGCGAG 1739

Qy 1621 CAGAAATGCTCAAGTATGATATATTAATTAAGTGTCAATGAGAAATAAACAAAGGTAT 1680
Db 1740 CAGAAATGCTCAAGTATGATATATTAATTAAGTGTCAATGAGAAATAAACAAAGGTAT 1799
Qy 1681 TCTATTTAGAAATGAGAAAAATCCTAACCATAAGATCCTGCAAAAGAAATCTGCTTTC 1740
Db 1800 TCTATTTAGAAATGAGAAAAATCCTAACCATAAGATCCTGCAAAAGAAATCTGCTTTC 1859
Qy 1741 AAAAGAAAGCTGAACCTATTAAGCAGCACTATTAAGCAATATGAGCAATTAATATTC 1800
Db 1860 AAAAGAAAGCTGAACCTATTAAGCAGCACTATTAAGCAATATGAGCAATTAATATTC 1919
Qy 1801 CACAATTTCAAAAGCACCTTAAAAAGATAGCGTAGAGAGAACTCTTACCAAGCATATT 1860
Db 1920 CACAATTTCAAAAGCACCTTAAAAAGATAGCGTAGAGAGAACTCTTACCAAGCATATT 1979
Qy 1861 CATGCCCTTGAACCTAGTACTAGTAAATCTTAAGCCCACTAATTTGATGAAATTCGA 1920
Db 1980 CATGCCCTTGAACCTAGTACTAGTAAATCTTAAGCCCACTAATTTGATGAAATTCGA 2039
Qy 1921 ATTGATAGTGTCTGACGATGAGATAAAGAAAAAAGTACAACTAATGCGAGTC 1980
Db 2040 ATTGATAGTGTCTGACGATGAGATAAAGAAAAAAGTACAACTAATGCGAGTC 2099
Qy 1981 AGGCACAGCAGAAACCTTACACTCATGGAAGTAAAGAACCTGCACTGAGAGCCAGAAAG 2040
Db 2100 AGGCACAGCAGAAACCTTACACTCATGGAAGTAAAGAACCTGCACTGAGAGCCAGAAAG 2159
Qy 2041 AGTAACAGCCCAATATGACAGACAAATGAATGAATGACATGATCTTCTCCAGAGCTG 2100
Db 2160 AGTAACAGCCCAATATGACAGACAAATGAATGAATGACATGATCTTCTCCAGAGCTG 2219
Qy 2101 AAGTTAACAATGACACCTGCTCTTTTACTAAGTTCATAATACAGTAATTAAGAA 2160
Db 2220 AAGTTAACAATGACACCTGCTCTTTTACTAAGTTCATAATACAGTAATTAAGAA 2279
Qy 2161 TTTGTCAATCTAGGCTTCCAGAGAGAAAAAAGAGAAATAGAAACAGTTAAAGTG 2220
Db 2280 TTTGTCAATCTAGGCTTCCAGAGAGAAAAAAGAGAAATAGAAACAGTTAAAGTG 2339
Qy 2221 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTGCANAAT 2280
Db 2340 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAAAGGTTTGCANAAT 2399
Qy 2281 GAAAGATCTGTAGAGATGACATATTTCAATGTTGATGTTGATGATGATGATGATGAT 2340
Db 2400 GAAAGATCTGTAGAGATGACATATTTCAATGTTGATGTTGATGATGATGATGATGATGAT 2459
Qy 2341 GAAAGTATCTGTTACTGGAAGTTAGACCTTAGGGAAGGCAAAAACAGAACCAATATA 2400
Db 2460 GAAAGTATCTGTTACTGGAAGTTAGACCTTAGGGAAGGCAAAAACAGAACCAATATA 2519
Qy 2401 TGTGTAGTCACTGTGACAGATTTTGAAGAACCCCAAGGACATATTCATGTTTCCAAA 2460
Db 2520 TGTGTAGTCACTGTGACAGATTTTGAAGAACCCCAAGGACATATTCATGTTTCCAAA 2579
Qy 2461 GATTAATAGAAATGACACAGAAAGCTTTAAGTATCATTTGGGGGCAATGAATTAACCAAGT 2520
Db 2580 GATTAATAGAAATGACACAGAAAGCTTTAAGTATCATTTGGGGGCAATGAATTAACCAAGT 2639
Qy 2521 CGGGAACACAGCATGGAATGGAAGAAAGTAACTTATGCTCATGTTTGCAGAAATACA 2580
Db 2640 CGGGAACACAGCATGGAATGGAAGAAAGTAACTTATGCTCATGTTTGCAGAAATACA 2699
Qy 2581 TTCAAGGTTTCAAGCCGACGATCTTGTCTGTTTTCATTAATCCAGGAAATGCAAGAG 2640
Db 2700 TTCAAGGTTTCAAGCCGACGATCTTGTCTGTTTTCATTAATCCAGGAAATGCAAGAG 2759
Qy 2641 GAATGTGACACATCTGCTGCCACATCTGGGGCTTTAAAGAAACAAAGTCAAAAGTCACT 2700
Db 2760 GAATGTGACACATCTGCTGCCACATCTGGGGCTTTAAAGAAACAAAGTCAAAAGTCACT 2819
```



QY 2701 TTGGAATGTGAACAAAAGAAATCAAGAAAGATAGCTTAATATCAAGCCCTGTA 2760  
 Db 2820 TTGGAATGTGAACAAAAGAAATCAAGAAAGATAGCTTAATATCAAGCCCTGTA 2879  
 QY 2761 CAGACATTAATATCACTGCAGGCTTCTGCTGTTGTTGTAAGAAATTAAGCAAGTTGAT 2820  
 Db 2880 CAGACATTAATATCACTGCAGGCTTCTGCTGTTGTTGTAAGAAATTAAGCAAGTTGAT 2939  
 QY 2821 AATGCCAAATGTATCAAGAGGAGCTAGGTTTGTCTATCATCTCAGTTCAGAGGC 2880  
 Db 2940 AATGCCAAATGTATCAAGAGGAGCTAGGTTTGTCTATCATCTCAGTTCAGAGGC 2999  
 QY 2881 AAGGAACCTGCATCACTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCGTATA 2940  
 Db 3000 AAGGAACCTGCATCACTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCGTATA 3059  
 QY 2941 CCACGACTTTTCCCATCAAGTCATTTGTTAAATCTAAATGTAAGAAAATCTGCTAGAG 3000  
 Db 3060 CCACGACTTTTCCCATCAAGTCATTTGTTAAATCTAAATGTAAGAAAATCTGCTAGAG 3119  
 QY 3001 GAAACCTTTGAGGAACATTCATGTCCATGTCGAAAGAGAAATGGGAAATGAGAACATTCGA 3060  
 Db 3120 GAAACCTTTGAGGAACATTCATGTCCATGTCGAAAGAGAAATGGGAAATGAGAACATTCGA 3179  
 QY 3061 AGTACAGTGAACAATTTAGCCCTAATTAACATTAGAGAAAATGTTTTTAAAGAGCCAGC 3120  
 Db 3180 AGTACAGTGAACAATTTAGCCCTAATTAACATTAGAGAAAATGTTTTTAAAGAGCCAGC 3239  
 QY 3121 TCAAGCAATTTAATGAGTAGTGTCCAGTACTAATGAAATGGGCTCCAGTATTAATGAA 3180  
 Db 3240 TCAAGCAATTTAATGAGTAGTGTCCAGTACTAATGAAATGGGCTCCAGTATTAATGAA 3299  
 QY 3181 ATAGGTTCCAGTATGAAACATTTCAAGAGACTAGTGAAGAAAGAGGAGCCAAATTTG 3240  
 Db 3300 ATAGGTTCCAGTATGAAACATTTCAAGAGACTAGTGAAGAAAGAGGAGCCAAATTTG 3359  
 QY 3241 AATGCTATGCTATGATTAAGGGGTTTGGCAACCTGAGGCTCTAATAACAAAGCTTCTGTA 3300  
 Db 3360 AATGCTATGCTATGATTAAGGGGTTTGGCAACCTGAGGCTCTAATAACAAAGCTTCTGTA 3419  
 QY 3301 AGTAATTTGATAGATCTCTGAAATTAAGAAAGCAAGATATGAAGAGTAGTTCAGACTGTT 3360  
 Db 3420 AGTAATTTGATAGATCTCTGAAATTAAGAAAGCAAGATATGAAGAGTAGTTCAGACTGTT 3479  
 QY 3361 AATACAGATTTCTCCATATCTGATTTGATGATTAAGAGCCCTATGGGAGAGT 3420  
 Db 3480 AATACAGATTTCTCCATATCTGATTTGATGATTAAGAGCCCTATGGGAGAGT 3539  
 QY 3421 CATGCACTTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAG 3480  
 Db 3540 CATGCACTTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAATTAAG 3599  
 QY 3481 GAGATATGATGTTTGTCTGAAATTAAGCAATTAAGAAAGTTCTGCTGTTTACCAAAAGC 3540  
 Db 3600 GAGATATGATGTTTGTCTGAAATTAAGCAATTAAGAAAGTTCTGCTGTTTACCAAAAGC 3659  
 QY 3541 GTCCAGAAAAGAGAGTTAGCAGAGAGTCTAGAGCCCTTACCAACCATACATTTGGCTCAG 3600  
 Db 3660 GTCCAGAAAAGAGAGTTAGCAGAGAGTCTAGAGCCCTTACCAACCATACATTTGGCTCAG 3719  
 QY 3601 GGTTCACGAAGAGGGCCCAAGAAATTAAGTCTCTCAGAGAGAACTTATCTAGTGAAGAT 3660  
 Db 3720 GGTTCACGAAGAGGGCCCAAGAAATTAAGTCTCTCAGAGAGAACTTATCTAGTGAAGAT 3779  
 QY 3661 GAAGAGCTTCCGCTTCCACACCTGTATATTTGGTAAGTAAACAATATACCTTCTCAG 3720  
 Db 3780 GAAGAGCTTCCGCTTCCACACCTGTATATTTGGTAAGTAAACAATATACCTTCTCAG 3839  
 QY 3721 TCTACTAGCATAGACACCTGTTGCTACGAGTGTCTGCTAAGAAACAGAGAGGAATTTA 3780  
 Db 3840 TCTACTAGCATAGACACCTGTTGCTACGAGTGTCTGCTAAGAAACAGAGAGGAATTTA 3899  
 QY 3781 TTATCATTTGAAGAAATGCTTAAATGACTGACGTACACAGGTAATATTTGGCAAGGACATCT 3840

Db 3900 TTATCATTTGAAGAAATFACCTTAATAGACTGCAGTAAACAGGTAATATTTGGCAAGGACATCT 3959  
 QY 3841 CAGGACATCACTTACTAGTGAGAAACAAATGTTCTGCTAGCTGTTTCTTCACAGTGC 3900  
 Db 3960 CAGGACATCACTTACTAGTGAGAAACAAATGTTCTGCTAGCTGTTTCTTCACAGTGC 4019  
 QY 3901 AGTGAATTTGGAACACTGACTGCAAAATACAAACCCAGATATCTTCTGATTTGTTCT 3960  
 Db 4020 AGTGAATTTGGAACACTGACTGCAAAATACAAACCCAGATATCTTCTGATTTGTTCT 4079  
 QY 3961 TCCAAACAAATGAGGATCACTGCTGAAAGCCAGGAGTGTGCTGAGTCAAGAAATTTG 4020  
 Db 4080 TCCAAACAAATGAGGATCACTGCTGAAAGCCAGGAGTGTGCTGAGTCAAGAAATTTG 4139  
 QY 4021 GTTTCAGATGATGAGAAAGAGAGAGCGGCTTGGAGAAATTAATCAAGAGAGCAAGC 4080  
 Db 4140 GTTTCAGATGATGAGAAAGAGAGAGCGGCTTGGAGAAATTAATCAAGAGAGCAAGC 4199  
 QY 4081 ATGGATTCAAACTTAGTGAAGAGCATCTGGGTTGAGTGAAGCAACGCTCTCGAA 4140  
 Db 4200 ATGGATTCAAACTTAGTGAAGAGCATCTGGGTTGAGTGAAGCAACGCTCTCGAA 4259  
 QY 4141 GACTGCTCAGGAGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATG 4200  
 Db 4260 GACTGCTCAGGAGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGGATACCATG 4319  
 QY 4201 CAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGTAAGTGAAGTGTGTTAGAACAG 4260  
 Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGTAAGTGAAGTGTGTTAGAACAG 4379  
 QY 4261 CATGGGAGCCAGCTTCTATACAGTACCTTCCATCATTAAGTACCTTCTGCTGAG 4320  
 Db 4380 CATGGGAGCCAGCTTCTATACAGTACCTTCCATCATTAAGTACCTTCTGCTGAG 4439  
 QY 4321 GACCTCGGAATCCAGAACCAACACATCAAGAAAAGCACTATTTAACTTCCAGAAAAGT 4380  
 Db 4440 GACCTCGGAATCCAGAACCAACACATCAAGAAAAGCACTATTTAACTTCCAGAAAAGT 4499  
 QY 4381 AGTGAATACCCTTAAGCCAGAAATCCAGAAAGCCCTTCTGCTCAAGATTTGAGGTGCT 4440  
 Db 4500 AGTGAATACCCTTAAGCCAGAAATCCAGAAAGCCCTTCTGCTCAAGATTTGAGGTGCT 4559  
 QY 4441 GCAGATAGTCTCCAGTAAATTAAGAAACAGGAGTGAAGAGTATCCCTCTCAAA 4500  
 Db 4560 GCAGATAGTCTCCAGTAAATTAAGAAACAGGAGTGAAGAGTATCCCTCTCAAA 4619  
 QY 4501 TGCCCATCATTAAGATGATAGTGTGATCATGACAGTGTGCTGAGAGTCTTCAGAAATGA 4560  
 Db 4620 TGCCCATCATTAAGATGATAGTGTGATCATGACAGTGTGCTGAGAGTCTTCAGAAATGA 4679  
 QY 4561 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAAGACACAGCTGGAA 4620  
 Db 4680 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAAGAGCTGGAA 4739  
 QY 4621 GAGTCTGGGCAACAGATTTGAGGAGAAACATTTACTTGGCAAGGCAAGATCTAGAGGGA 4680  
 Db 4740 GAGTCTGGGCAACAGATTTGAGGAGAAACATTTACTTGGCAAGGCAAGATCTAGAGGGA 4799  
 QY 4681 ACCCCTTACCTGGAATCTGGAATCAAGCTTCTCTGATGACCCCTGAATCTGATCTTCT 4740  
 Db 4800 ACCCCTTACCTGGAATCTGGAATCAAGCTTCTCTGATGACCCCTGAATCTGATCTTCT 4859  
 QY 4741 GAAGACAGAGCCCAAGCTCAGTGTGTTGGCAACATACATCTTCAACCTCTGCATTG 4800  
 Db 4860 GAAGACAGAGCCCAAGCTCAGTGTGTTGGCAACATACATCTTCAACCTCTGCATTG 4919  
 QY 4801 AAGGTTCCCAATTTGAAGTTGAGAAATCTGCCAAGTCCAGCTGCTGCTACTACT 4860  
 Db 4920 AAGGTTCCCAATTTGAAGTTGAGAAATCTGCCAAGTCCAGCTGCTGCTACTACT 4979  
 QY 4861 GATACGTGCTGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGAGAGCAGAAATTGACA 4920

Dd		4980	GATACCTGCGGTATAAATGCATAGSAAACAAGAAGTGGACAGGAGGAACCAGAAATTGACA	5039
Oy		4921	GCTTCACAGAAAAGGGTCAACAAAAAGATGCCATGTTGGTGCTGTGCGCTGACCCACAGAA	4980
Dd		5040	GCTTCACAGAAAAGGGTCAACAAAAAGATGCCATGTTGGTGCTGTGCGCTGACCCACAGAA	5099
Oy		4981	GAATTATAGCCTGCTACAGATTTCCAGAAAACCCCATCACACTTTAACATCTAATT	5040
Dd		5100	GAATTTATGCTGCTGTACAGTTTTTCCAGAAAACCCCATCACACTTTAACATCTAATT	5159
Oy		5041	ACTGAAGACACTCATCTGTTGTTATVGA AAAACAGATGCTGAGTTTGCTGTGAACGACA	5100
Dd		5160	ACTGAAGACACTCATCTGTTGTTATVGA AAAACAGATGCTGAGTTTGCTGTGAACGACA	5219
Oy		5101	CTGA AATATTTTCTCTGGAATGCGGGAGAGAAAATGGTAGTACTATTCTGGCTGACC	5160
Dd		5220	CTGA AATATTTTCTCTGGAATGCGGGAGAGAAAATGGTAGTACTATTCTGGCTGACC	5279
Oy		5161	CAGTCTATTAAAGAAAAGAAAATGCTGTAATGAGCATGATTTTGAAGTCAGAGGAGATGTG	5220
Dd		5280	CAGTCTATTAAAGAAAAGAAAATGCTGTAATGAGCATGATTTTGAAGTCAGAGGAGATGTG	5339
Oy		5221	GTCATATGGAAGAACCCACCAAAGTCCAAAGCGAGCAGAGAAATCCCAGACAGAAAATGC	5280
Dd		5340	GTCATATGGAAGAACCCACCAAAGTCCAAAGCGAGCAGAGAAATCCCAGACAGAAAATGC	5399
Oy		5281	TTCAGGGGGCTAGAAATCTGTGTCATG GGCCTTCACCA C ATGAGCCACAGATCAACTG	5340
Dd		5400	TTCAGGGGGCTAGAAATCTGTGTCATG GGCCTTCACCA C ATGAGCCACAGATCAACTG	5459
Oy		5341	GAATGATGTATACAGCTGTGTGTGCTCTGTGTGTAAGAGAGCTTTCATCTACCCCTT	5400
Dd		5460	GAATGATGTATACAGCTGTGTGTGCTCTGTGTGTAAGAGAGCTTTCATCTACCCCTT	5519
Oy		5401	GGCAGAGTGTCCACCAATTTGTGGTTGTGGACAGATGCTGTGAGACAGACAATGGC	5460
Dd		5520	GGCAGAGTGTCCACCAATTTGTGGTTGTGGACAGATGCTGTGAGACAGACAATGGC	5579
Oy		5461	TTCATGCAATTTGGGACAGATGTGTGAGGCACCTGTGTGACCCGAGAGTGGTGTGAGC	5520
Dd		5580	TTCATGCAATTTGGGACAGATGTGTGAGGCACCTGTGTGACCCGAGAGTGGTGTGAGC	5639
Oy		5521	AGTGAAGAC TCTACCACTGSCCAGAGCTGAGACAC TCACTGATACACCAATCCCCAC	5580
Dd		5640	AGTGAAGAC TCTACCACTGSCCAGAGCTGAGACAC TCACTGATACACCAATCCCCAC	5699
Oy		5581	AGCCACTATC 5589         	
Dd		5700	AGCCACTATC 5708	
RESULT 11				
AAV46462	ID	AAV46462 standard; cDNA; 5711 BP.		
XX	AC	AAV46462;		
XX	DT	18-NOV-1998 (first entry)		
DE	DE	Human BRCA1 omi2 polymorphism #5 cDNA.		
KW	KW	BRCA1: omi2; human: breast and ovarian cancer predisposing gene:		
KM	KM	polymorphism: susceptibility; anti-oncogene; tumour suppressor:		
XX	OS	Homo sapiens.		
FH	FT	Key location/Qualifiers		
FT	CD	120..5711		
FT	FT	/+tag= a		
FT	FT	/product= "BRCA1 omi2 protein"		
FT	FT	variation 3667		
FT	FT	/+tag= b		

Query Match	Best Local Similarity	Score	DB	Length
Matches 5587; Conservative 1; Mismatches 1; Indels 0; Gaps 0	100.0%;	5587;	DB 19;	Length 5711;
1	ATGGATTATCTCTCTCTCGCGTTGAGAGACAAATGTCATTATGCTATGACGAAA	60		
120	ATGATATTATCTCTCTCTCGCGTTGAGAGACAAATGTCATTATGCTATGACGAAA	179		
61	ATCTTAGAGTGTCCATCTGTCTGAGATGATCAAGAACCTGTCTCCAAAGTGTGAC	120		
180	ATCTTAGAGTGTCCATCTGTCTGAGATGATCAAGAACCTGTCTCCAAAGTGTGAC	239		
121	CACATATTTTGGCAAAATTTTGGCATGCGTGAACCTTCACACGAGAAAGGCGCTTCACAG	180		
240	CACATATTTTGGCAAAATTTTGGCATGCGTGAACCTTCACACGAGAAAGGCGCTTCACAG	299		
181	TGTCCTTTATGTAAAGATGATTAACCAAAAGGAGACCTTCACAGAAATGAGATTAGT	240		
300	TGTCCTTTATGTAAAGATGATTAACCAAAAGGAGACCTTCACAGAAATGAGATTAGT	359		
241	CAACTTGTGAGAGCTATTGAAATCATTTGTGCTTTACGCTTGACACAGCTTTTGAG	300		
360	CAACTTGTGAGAGCTATTGAAATCATTTGTGCTTTACGCTTGACACAGCTTTTGAG	419		
301	TATGCAAAACGCAATATTTTGGCAAAAGGAAATATACCTCTCGTAACATTCAAAGAT	360		
420	TATGCAAAACGCAATATTTTGGCAAAAGGAAATATACCTCTCGTAACATTCAAAGAT	479		
361	GAAATTCTATCATCCAAAGTATGGGCTACAGAAACCGTCCAAAGACCTTACAGAGT	420		
480	GAAATTCTATCATCCAAAGTATGGGCTACAGAAACCGTCCAAAGACCTTACAGAGT	539		

QY 421 GAAACCGAAATCCTTCCTGAGGAAACCAAGTCTAGTGTCCAACTCTTAACCTTGA 480  
DB 540 GAACCCGAAATCCTTCCTGAGGAAACCAAGTCTAGTGTCCAACTCTTAACCTTGA 599  
QY 481 ACTGTAGAACTGTGAGGACAAAGCGGATACAACTCAAAAAGCTGTCTACATT 540  
DB 600 ACTGTAGAACTGTGAGGACAAAGCGGATACAACTCAAAAAGCTGTCTACATT 659  
QY 541 GAATTGGATCTGATTCCTTGAAGATACCGTTAATAGCACTTATTGACAGTGGGA 600  
DB 660 GAATTGGATCTGATTCCTTGAAGATACCGTTAATAGCACTTATTGACAGTGGGA 719  
QY 601 GATCAAGAAATTTTACAAATCACCCCTCAAGAACCGGATGAAATCACTTTGATTC 660  
DB 720 GATCAAGAAATTTTACAAATCACCCCTCAAGAACCGGATGAAATCACTTTGATTC 779  
QY 661 GCAAAAAAGCTGCTGTGATTTTCTGAGAGCGGATGTAACTAACTGACATCA 720  
DB 780 GCAAAAAAGCTGCTGTGATTTTCTGAGAGCGGATGTAACTAACTGACATCA 839  
QY 721 CCCAGTATATGATTTGAACACACTGAGAGCGTGCAGTGAAGCGCTCCAGAAAG 780  
DB 840 CCCAGTATATGATTTGAACACACTGAGAGCGTGCAGTGAAGCGCTCCAGAAAG 899  
QY 781 TATCAGGATGATTCCTGTTTCAAACTTGCATGTGAGCCATGTGGCAAAATCTCATGCC 840  
DB 900 TATCAGGATGATTCCTGTTTCAAACTTGCATGTGAGCCATGTGGCAAAATCTCATGCC 959  
QY 841 AGCTCATTTACAGATGGAACAGCAGTTTATTAATCTCTAAAGACAGATGAATGA 900  
DB 960 AGCTCATTTACAGATGGAACAGCAGTTTATTAATCTCTAAAGACAGATGAATGA 1019  
QY 901 AAGGCTGAATTCGTATATAAGCAACAGCTGGCTTACAGAGGCCAATACAGA 960  
DB 1020 AAGGCTGAATTCGTATATAAGCAACAGCTGGCTTACAGAGGCCAATACAGA 1079  
QY 961 TGGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGATCCAGCAGAGAAAAAGGTA 1020  
DB 1080 TGGGCTGGAAGTAAAGAAACATGTAATGATAGGCGGATCCAGCAGAGAAAAAGGTA 1139  
QY 1021 GATCTGAATCTGATCCCTGCTGTGAGAGAAAAAGATGAAATACAGAAATCTCATGCC 1080  
DB 1140 GATCTGAATCTGATCCCTGCTGTGAGAGAAAAAGATGAAATACAGAAATCTCATGCC 1199  
QY 1081 TCAGAGATCTAGAGATGTAAGATGTCTTGATTAACACTAAATAGCAGATTCAG 1140  
DB 1200 TCAGAGATCTAGAGATGTAAGATGTCTTGATTAACACTAAATAGCAGATTCAG 1259  
QY 1141 AAGATTAATGATGTTTTTCCAGAGTGAATGATGTTAGTCTGTGATGCTCATGAT 1200  
DB 1260 AAGATTAATGATGTTTTTCCAGAGTGAATGATGTTAGTCTGTGATGCTCATGAT 1319  
QY 1201 GGGGAGCTGAATCAAAATGCCAAAGTACTGATGATTTGAGCCTTCTAAATGAGTAGAT 1260  
DB 1320 GGGGAGCTGAATCAAAATGCCAAAGTACTGATGATTTGAGCCTTCTAAATGAGTAGAT 1379  
QY 1261 GAATTTCTGCTTCTCAGAGAAATAGACTTACTGGCCAGTACTCTCATGAGGCTTTA 1320  
DB 1380 GAATTTCTGCTTCTCAGAGAAATAGACTTACTGGCCAGTACTCTCATGAGGCTTTA 1439  
QY 1321 ATATGTAAGTGAAGAGTTCATCCAAATCAGTAGAGATATATTGAGAGCAAAATA 1380  
DB 1440 ATATGTAAGTGAAGAGTTCATCCAAATCAGTAGAGATATATTGAGAGCAAAATA 1499  
QY 1381 TTTGGGAAAACTATGGAAGAGGAGCTCCCAACTTAAGCCATGTAACTGAAAT 1440  
DB 1500 TTTGGGAAAACTATGGAAGAGGAGCTCCCAACTTAAGCCATGTAACTGAAAT 1559  
QY 1441 CTAAATTTAGAGCATTTGTTACTGAGCCAGATATATACAGAGGCTCCCTCCACAAAT 1500  
DB 1560 CTAAATTTAGAGCATTTGTTACTGAGCCAGATATATACAGAGGCTCCCTCCACAAAT 1619

QY 1501 AAATTAAGCGTAAAGAGACCTACATCAGGCCCTTCATCTGAGATTTTATCAAGAA 1560  
DB 1620 AAATTAAGCGTAAAGAGACCTACATCAGGCCCTTCATCTGAGATTTTATCAAGAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAGACTCCTGGAATGATTAATCAGGAACTAAACCAAGGAG 1620  
DB 1680 GCAGATTTGGCAGTTCAAAGACTCCTGGAATGATTAATCAGGAACTAAACCAAGGAG 1739  
QY 1621 CAGATTTGGTCAAGTGAATATTTACTAATAGTGTGATGATGAATTAACAAAGGTGAT 1680  
DB 1740 CAGATTTGGTCAAGTGAATATTTACTAATAGTGTGATGATGAATTAACAAAGGTGAT 1799  
QY 1681 TCTATTGAGATGAGAAAAATCCTAACCCATAGATCACTCGAAAAAGATCTGTTTC 1740  
DB 1800 TCTATTGAGATGAGAAAAATCCTAACCCATAGATCACTCGAAAAAGATCTGTTTC 1859  
QY 1741 AAAACGAAAGCTGACCTTATAGCAGAGATTAAGCAATTTGGAACCTCGAATTAATATC 1800  
DB 1860 AAAACGAAAGCTGACCTTATAGCAGAGATTAAGCAATTTGGAACCTCGAATTAATATC 1919  
QY 1801 CACAAATCAAAAGCACCTTAATAAAGATAGGCTGAGAGAGTCTTCTACAGGATATT 1860  
DB 1920 CACAAATCAAAAGCACCTTAATAAAGATAGGCTGAGAGAGTCTTCTACAGGATATT 1979  
QY 1861 CATGCGCTGACACTAGTAGTCACTAGAAATCTAAGCCACCTAATTGTACTGAAATTC 1920  
DB 1980 CATGCGCTGACACTAGTAGTCACTAGAAATCTAAGCCACCTAATTGTACTGAAATTC 2039  
QY 1921 ATTGATTTGTTTCTTACAGTGAAGATTAAGCAAAAGATTAACCAAAATGCGAGTC 1980  
DB 2040 ATTGATTTGTTTCTTACAGTGAAGATTAAGCAAAAGATTAACCAAAATGCGAGTC 2099  
QY 1981 AGGCACAGCAGAAAGCTACAGCTCATGGAAGTAAAGAACCTGACATGAGCCAGAAAG 2040  
DB 2100 AGGCACAGCAGAAAGCTACAGCTCATGGAAGTAAAGAACCTGACATGAGCCAGAAAG 2159  
QY 2041 AGTAAACGCAAAATGAAACAGACAGTAAAGAAAGATGACAGTACTTCTCCAGAGCTG 2100  
DB 2160 AGTAAACGCAAAATGAAACAGACAGTAAAGAAAGATGACAGTACTTCTCCAGAGCTG 2219  
QY 2101 AAGTTAACAATGCGACCTGCTTTTACTAATGTTCAATACAGTGAATTAAGAA 2160  
DB 2220 AAGTTAACAATGCGACCTGCTTTTACTAATGTTCAATACAGTGAATTAAGAA 2279  
QY 2161 TTTGTCATCTGACCTTCCAGAGAAAGAAAGAAAGAAAGTAAAGG 2220  
DB 2280 TTTGTCATCTGACCTTCCAGAGAAAGAAAGAAAGAAAGTAAAGG 2339  
QY 2221 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTGGCAACT 2280  
DB 2340 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTTGGCAACT 2399  
QY 2281 GAAAGATCTGTAGAGATGACAGTATTTCAATGCTGTGACTGATTAATGCGACTCAG 2340  
DB 2400 GAAAGATCTGTAGAGATGACAGTATTTCAATGCTGTGACTGATTAATGCGACTCAG 2459  
QY 2341 GAAAGATCTGTTACTGGAAGTTAGCAGCTGTGGGAAGGCAAAACAGAACCAATAAA 2400  
DB 2460 GAAAGATCTGTTACTGGAAGTTAGCAGCTGTGGGAAGGCAAAACAGAACCAATAAA 2519  
QY 2401 TGTGTAGTCACTGCTCAGCATTTGAAACCCCAAGGAGTAAATTCATGCTTTC 2460  
DB 2520 TGTGTAGTCACTGCTCAGCATTTGAAACCCCAAGGAGTAAATTCATGCTTTC 2579  
QY 2461 GATTAATTAAGATGACACAGAGCTTTAATATTCATTTGGAGATGAATTAACACAGT 2520  
DB 2580 GATTAATTAAGATGACACAGAGCTTTAATATTCATTTGGAGATGAATTAACACAGT 2639  
QY 2521 CCGGAAACAAAGCTAGAAATGGAAGAAAGTGAATGATGCTGATGATTTGCAATPAC 2580  
DB 2640 CCGGAAACAAAGCTAGAAATGGAAGAAAGTGAATGATGCTGATGATTTGCAATPAC 2699  
QY 2581 TTCAGGTTTCAAGCGCAGTCAATTTGCTGTTTCAAAATCCAGAAATGCGAAGAG 2640

Db 2700 TTCAAGGTTTCAAAAGGCCAGTCATTTGCTCTTTTCAATCCAGAAATGCAGAAAG 2759  
Qy 2641 GAATGCGCAACATCTCTGCCCCACTGCGGTCTCTTAAAGAAACAAAGTCCAAAAGTCAC 2700  
Db 2760 GAATGCGCAACATCTCTGCCCCACTGCGGTCTCTTAAAGAAACAAAGTCCAAAAGTCAC 2819  
Qy 2701 TTGTAATGTAACAAAGAAAGAAATCAAGAAAGAAATGAGTCTATATCAAGCCGTGA 2760  
Db 2820 TTGTAATGTAACAAAGAAAGAAATCAAGAAAGAAATGAGTCTATATCAAGCCGTGA 2879  
Qy 2761 CAGACAGTTAATATCACTGAGGCTTCTGTTGGTTCAGAAAGATAAGCCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCACTGAGGCTTCTGTTGGTTCAGAAAGATAAGCCAGTTGAT 2939  
Qy 2821 AATGCCAAATAGTATCAAGAGAGGCTAGAGTTTGTCTATCATCTGCTCAGAGGC 2880  
Db 2940 AATGCCAAATAGTATCAAGAGAGGCTAGAGTTTGTCTATCATCTGCTCAGAGGC 2999  
Qy 2881 AAGCAAACTGAGCTCTTACTCCAAATAAACATGAGCTTTTACAAAACCCATATCGTATA 2940  
Db 3000 AAGCAAACTGAGCTCTTACTCCAAATAAACATGAGCTTTTACAAAACCCATATCGTATA 3059  
Qy 2941 CCACCACTTTTCCCATCAAGTCATTTGTTAAACTAATGTAAGAAAATCTGCTAGAG 3000  
Db 3060 CCACCACTTTTCCCATCAAGTCATTTGTTAAACTAATGTAAGAAAATCTGCTAGAG 3119  
Qy 3001 GAAAACCTTGAGGAACATTCATGTCACCTGAAAGAAATGGAAATGGAACATTTCCA 3060  
Db 3120 GAAAACCTTGAGGAACATTCATGTCACCTGAAAGAAATGGAAATGGAACATTTCCA 3179  
Qy 3061 AGTACAGTGAAGCAATTAAGCCGTAAATCAATTAGAAATGTTTAAAGAGCCAGC 3120  
Db 3180 AGTACAGTGAAGCAATTAAGCCGTAAATCAATTAGAAATGTTTAAAGAGCCAGC 3239  
Qy 3121 TCAAGCAATTAATGAAAGTAGTTCACATCTAATGAAAGTGGGCTCCAGTATTAATGAA 3180  
Db 3240 TCAAGCAATTAATGAAAGTAGTTCACATCTAATGAAAGTGGGCTCCAGTATTAATGAA 3299  
Qy 3181 ATAGGTTCCAGTATGAAACATTTCAAGAGAACTGTAAGAAAGAGAGGCCCAAAATG 3240  
Db 3300 ATAGGTTCCAGTATGAAACATTTCAAGAGAACTGTAAGAAAGAGAGGCCCAAAATG 3359  
Qy 3241 AATGCTATGCTTAGATTAGGGTTTTGCAACCTGAGGTCTATTAACAAAGTCTTCTGGA 3300  
Db 3360 AATGCTATGCTTAGATTAGGGTTTTGCAACCTGAGGTCTATTAACAAAGTCTTCTGGA 3419  
Qy 3301 AGTAATTTGTAAGCATCTGAAATTAAGAAAGCAAGAAATGAAAGAGTAGTTCAGACTGT 3360  
Db 3420 AGTAATTTGTAAGCATCTGAAATTAAGAAAGCAAGAAATGAAAGAGTAGTTCAGACTGT 3479  
Qy 3361 AATACAGATTTCTCCCATATGATGATGAGATTAAGTAAACAGCTATGGAAGTATAG 3420  
Db 3480 AATACAGATTTCTCCCATATGATGATGAGATTAAGTAAACAGCTATGGAAGTATAG 3539  
Qy 3421 CATGATCTCAGAGTTGTTCTAGAGACCTGATGACCTTAGATGATGATGATGATGATGAT 3480  
Db 3540 CATGATCTCAGAGTTGTTCTAGAGACCTGATGACCTTAGATGATGATGATGATGATGAT 3599  
Qy 3481 GAAAGTACTAGTTTGTGCTGAAATGACATTAAGAAAGTTCGCTGTTTTTACCAAAAGC 3540  
Db 3600 GAAAGTACTAGTTTGTGCTGAAATGACATTAAGAAAGTTCGCTGTTTTTACCAAAAGC 3659  
Qy 3541 GTCCGAAAGAGAGAGCTTAGAGAGTCTAGAGCTTACACCATACATTTGGCTGAG 3600  
Db 3660 GTCCGAAAGAGAGAGCTTAGAGAGTCTAGAGCTTACACCATACATTTGGCTGAG 3719  
Qy 3601 GGTACCGAAGAGAGGCGCAAGAAATTAAGAGTCTCTAGAGAGAACTATATAGTGGAGAT 3660  
Db 3720 GGTACCGAAGAGAGGCGCAAGAAATTAAGAGTCTCTAGAGAGAACTATATAGTGGAGAT 3779  
Qy 3661 GAAGAGCTTCCCTGCTTCCAAACATTTGTTATTTGTAAGTAACAATATACCTTCAG 3720  
Db 3780 GAAGAGCTTCCCTGCTTCCAAACATTTGTTATTTGTAAGTAACAATATACCTTCAG 3839  
Qy 3721 TCTACTAGGATATGACACCGGTGCTACCCGAGTCTGCTGCTAACAACAGAGAGATTTA 3780  
Db 3840 TCTACTAGGATATGACACCGGTGCTACCCGAGTCTGCTGCTAACAACAGAGAGATTTA 3899  
Qy 3781 TTATCATTTAGAAATAGCTTTAAATGACTGACATAACAGAGTAAATTTGCAAAAGCATCT 3840  
Db 3900 TTATCATTTAGAAATAGCTTTAAATGACTGACATAACAGAGTAAATTTGCAAAAGCATCT 3959  
Qy 3841 CAGAAACATACCTTAGTAGAGAAACAAATGTTCTGCTAGCTGTTTTTCTCACAGTGC 3900  
Db 3960 CAGAAACATACCTTAGTAGAGAAACAAATGTTCTGCTAGCTGTTTTTCTCACAGTGC 4019  
Qy 3901 AGTGAATTTGGAAGACTTGAACCTGCAAAATCAAAACCCAGAGATCTTTCTTATGAGTCT 3960  
Db 4020 AGTGAATTTGGAAGACTTGAACCTGCAAAATCAAAACCCAGAGATCTTTCTTATGAGTCT 4079  
Qy 3961 TCCAAACAAATGAGGATGATGTAAGCCAGGAGTTGGTCTGAGTGACAGAAATTTG 4020  
Db 4080 TCCAAACAAATGAGGATGATGTAAGCCAGGAGTTGGTCTGAGTGACAGAAATTTG 4139  
Qy 4021 GTTTCAGATGATGAAGAAAGAGAAAGGCTTTGGAAGAAATATCAAGAAAGCAAGC 4080  
Db 4140 GTTTCAGATGATGAAGAAAGAGAAAGGCTTTGGAAGAAATATCAAGAAAGCAAGC 4199  
Qy 4081 ATGGAATTTCAACTTACGTAAGCAGATCTGGGTGAGAGTAACAAGGCTCTGAA 4140  
Db 4200 ATGGAATTTCAACTTACGTAAGCAGATCTGGGTGAGAGTAACAAGGCTCTGAA 4259  
Qy 4141 GACTGCTCAGGCGTATCTCTCAGAGTACATTTTAAACACTCAGCAGAGGATACCATG 4200  
Db 4260 GACTGCTCAGGCGTATCTCTCAGAGTACATTTTAAACACTCAGCAGAGGATACCATG 4319  
Qy 4201 CAACATTAACCTGATTAAGAGTCCAGCAGAGAAATGGCTGAACACTGATGAGACAG 4260  
Db 4320 CAACATTAACCTGATTAAGAGTCCAGCAGAGAAATGGCTGAACACTGATGAGACAG 4379  
Qy 4261 CATGGAGGAGGAGCTTTCAACAGTACCTTCATCATATAGTACCTTTCGCCCTGAG 4320  
Db 4380 CATGGAGGAGGAGCTTTCAACAGTACCTTCATCATATAGTACCTTTCGCCCTGAG 4439  
Qy 4321 GACCTGCGAAATTCAGAAACAAAGCAATCAGAAAGAGTAACTTCAACAGAAAGT 4380  
Db 4440 GACCTGCGAAATTCAGAAACAAAGCAATCAGAAAGAGTAACTTCAACAGAAAGT 4499  
Qy 4381 AGTGAATACCTTATTAAGCCAGATCCAGAGGCTTTTCTGCTGACAGTTCGAGTGTCT 4440  
Db 4500 AGTGAATACCTTATTAAGCCAGATCCAGAGGCTTTTCTGCTGACAGTTCGAGTGTCT 4559  
Qy 4441 GCAGATAGTTTACAGTAAATTAAGAAAGCAGAGTGAAGAGTCAATCCCTTCTAAA 4500  
Db 4560 GCAGATAGTTTACAGTAAATTAAGAAAGCAGAGTGAAGAGTCAATCCCTTCTAAA 4619  
Qy 4501 TGCCTCATTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
Db 4620 TGCCTCATTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4679  
Qy 4561 AACTACCATTCAGAGAGAGGCTCAATTAAGTTGTTGATGAGAGAGCAACAGCTGAAA 4620  
Db 4680 AACTACCATTCAGAGAGAGGCTCAATTAAGTTGTTGATGAGAGAGCAACAGCTGAAA 4739  
Qy 4621 GAGTCTGGGCGACAGAGTTTGAAGAAACATCTTACCTGCAAGGCAAGATCTGAGGGA 4680  
Db 4740 GAGTCTGGGCGACAGAGTTTGAAGAAACATCTTACCTGCAAGGCAAGATCTGAGGGA 4799  
Qy 4681 ACCCTTACCTGGAATTTGGAATGAGCTCTTCTCTGATGAGACCTGATGATGATGATGAT 4740  
Db 4800 ACCCTTACCTGGAATTTGGAATGAGCTCTTCTCTGATGAGACCTGATGATGATGATGAT 4859  
Qy 4741 GAAGACAGAGCCCGAGAGTCAAGTCTGTTGGAACATACATCTTCAACCTTCGATTTG 4800  
Db 4860 GAAGACAGAGCCCGAGAGTCAAGTCTGTTGGAACATACATCTTCAACCTTCGATTTG 4919

```

QY 4801 AAGTCCCAATTTGAAGTTCGAGATCTGCCAGATCTCCTGCTACTACT 4860
    |||||||
Db 4920 AAGTCCCAATTTGAAGTTCGAGATCTGCCAGATCTCCTGCTACTACT 4879
QY 4861 GATACGCTGGGTATATGCAATGGAAGAAAGTGTGACGAGGAGAGCAATGACA 4920
    |||||||
Db 4980 GATACGCTGGGTATATGCAATGGAAGAAAGTGTGACGAGGAGAGCAATGACA 5039
QY 4921 GCTTCAACAGAAAGGCTCAACAAAGATGTCATGGTGTGGCTGACCCAGAA 4980
    |||||||
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGATGTCATGGTGTGGCTGACCCAGAA 5099
QY 4981 GAATTTATGCTGCTGATACAGTTTGGCAGAAACACACATCACTTAATTAATT 5040
    |||||||
Db 5100 GAATTTATGCTGCTGATACAGTTTGGCAGAAACACACATCACTTAATTAATT 5159
QY 5041 ACTGAAGAGACTACTGATGTTGTATGAANAACAGATGCTGATGTTGTGTAACGACA 5100
    |||||||
Db 5160 ACTGAAGAGACTACTGATGTTGTATGAANAACAGATGCTGATGTTGTGTAACGACA 5219
QY 5101 CTGAATATTTTCTAGGAATTTGGGGAGAAATGGGTAGTTAGCTATTTCTGGTGACC 5160
    |||||||
Db 5220 CTGAATATTTTCTAGGAATTTGGGGAGAAATGGGTAGTTAGCTATTTCTGGTGACC 5279
QY 5161 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGACATGATTTTGAAGTCAGAGAGATGTG 5220
    |||||||
Db 5280 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGACATGATTTTGAAGTCAGAGAGATGTG 5339
QY 5221 GTCAATGGAAGAAACCAACCAAGTCCAAAGCAGCAAGAGATCCCAAGCAAGAAAGATC 5280
    |||||||
Db 5340 GTCAATGGAAGAAACCAACCAAGTCCAAAGCAGCAAGAGATCCCAAGCAAGAAAGATC 5399
QY 5281 TTCAGGGGGCTAGAAATCTGTTCTATGAGGCCCTTACCAACATGCCCAGATCAACTG 5340
    |||||||
Db 5400 TTCAGGGGGCTAGAAATCTGTTCTATGAGGCCCTTACCAACATGCCCAGATCAACTG 5459
QY 5341 GAATGATGCTACAGTGTGTGCTTCTGTGTGTAAGAGAGCTTTCATCATTCACCCCT 5400
    |||||||
Db 5460 GAATGATGCTACAGTGTGTGCTTCTGTGTGTAAGAGAGCTTTCATCATTCACCCCT 5519
QY 5401 GGACAGAGTGTCCACCAATTTGTTGTGTCACGACAGATCCCTGACAGAGCAATGCC 5460
    |||||||
Db 5520 GGACAGAGTGTCCACCAATTTGTTGTGTCACGACAGATCCCTGACAGAGCAATGCC 5579
QY 5461 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGGTGACCGAGAGTGGGTGGAC 5520
    |||||||
Db 5580 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGGTGACCGAGAGTGGGTGGAC 5639
QY 5521 AGTGTAGCACTTACCAAGTCCAGAGAGTGGACACCTACTGATACCCAGATCCCCAC 5580
    |||||||
Db 5640 AGTGTAGCACTTACCAAGTCCAGAGAGTGGACACCTACTGATACCCAGATCCCCAC 5699
QY 5581 AGCCACTAC 5589
    |||||||
Db 5700 AGCCACTAC 5708

```

RESULT 12

AAT84840 standard; cDNA to mRNA; 5712 BP.

AAT84840;

21-FEB-1998 (first entry)

Human breast and ovarian cancer susceptibility gene BRCA1.

BRCA1; breast cancer; ovarian cancer; human;

tumour suppressor gene; gene therapy; receptor; ds.

Homo sapiens.

OS

XX

```

FH Key Location/Qualifiers
FT CDS 120..5711
    /*tag= a
    /transl_except= (pos:1482..1484, aa:Asp)
    /transl_except= (pos:1581..1583, aa:Ser)
    /transl_except= (pos:1899..1901, aa:Glu)
    /transl_except= (pos:1902..1904, aa:Leu)
    /transl_except= (pos:1905..1907, aa:Glu)
    /transl_except= (pos:1908..1910, aa:Leu)
    /transl_except= (pos:1911..1913, aa:Asn)
    /transl_except= (pos:1914..1916, aa:Ile)
    /transl_except= (pos:1917..1919, aa:Met)
    /transl_except= (pos:4476..4478, aa:Val)
    /transl_except= (pos:4479..4481, aa:Leu)
    /transl_except= (pos:4482..4484, aa:Gln)
    /transl_except= (pos:4689..4691, aa:Pro)

```

MO9730108-A1.

21-AUG-1997.

19-FEB-1997; 97MO-US03340.

20-FEB-1996; 96US-0603753.

(UYVA-) UNIV VANDERBILT.  
(UNIM ) UNIV WASHINGTON.Claire-King M, Holt JT, Jensen RA, Jetton TL, Page DL;  
Robinson-Benlion CL, Szabo CI, Thompson ME;

MPI; 1997-434733/40.

P-PSDB; AAW23286.

BRCA1 and BRCA2 tumour suppressor gene products - useful to inhibit breast and ovarian cancer cell growth and tumorigenesis, or treat gene linked hereditary or sporadic ovarian or breast cancer

Claim 13: Page 54-63; 148pp; English.

This sequence comprises a full-length BRCA1 cDNA. Genetic analysis of familial and ovarian cancer indicates that BRCA1 is a tumour suppressor gene. It encodes a 190 kDa protein (see AAW23286) that is an inhibitor of the growth and proliferation of human breast and ovarian cancer cells. DNA encoding the BRCA1 protein can thus be used in gene therapy methods for the treatment of breast and ovarian cancers. A purified BRCA1 protein can also be used to treat these cancers and, since it is secreted, can be used to identify the BRCA1 receptor and hence to identify BRCA1 protein-mimetic agents which act on the receptor for use in breast and ovarian cancer treatment. The BRCA2 gene (see AAT84841) and BRCA2 protein (see AAW23287) have also been characterised. Methods are claimed for the isolation of BRCA1 or BRCA2 receptors, and for treating or preventing breast and (gene-linked hereditary or sporadic) ovarian cancers using BRCA1 and BRCA2 proteins and genes.

Sequence 5712 BP; 1956 A; 1099 C; 1274 G; 1383 T; 0 other;

Query Match 99.9%; Score 5585.8; DB 18; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ATGGATTATCTGCTCTGCGGTTGAAGAGTCAAAAATGCTATTATGTCAGAGAAA 60
    |||||||
Db 120 ATGGATTATCTGCTCTGCGGTTGAAGAGTCAAAAATGCTATTATGTCAGAGAAA 179
QY 61 ATCTTAGAGTGTCCACTCTCTGAGATTGATCAAGGAACCTCTCTCCAGAGTGTGAC 120
    |||||||
Db 180 ATCTTAGAGTGTCCACTCTCTGAGATTGATCAAGGAACCTCTCTCCAGAGTGTGAC 239
QY 121 CACATATTTTGCAAAATTTTGCAATGCTGAACCTTCTCAACAGAGAAAGGCGCTTCACAG 180
    |||||||

```

D 240 CACATATTTTGCMAATTTTGCATGCTGAAACCTTCTCAACCAGAAAGGGGCTTTCAG 299  
Q 181 TGCCTTATGTAGATGATATACCAAAAGAGGCTTACAGAAAGTACAGATTAGT 240  
D 300 TGTCTTTTATGTAGATGATATACCAAAAGAGGCTTACAGAAAGTACAGATTAGT 359  
Q 241 CAACCTGTGAGAGCTATTTGAAATCATTTTGTCTTTTCAGCTTGACAGAGTTTGGAG 300  
D 360 CAACCTGTGAGAGCTATTTGAAATCATTTTGTCTTTTCAGCTTGACAGAGTTTGGAG 419  
Q 301 TATGCAACAGCTATTAATTTTGCMAAAAGGAAATTAACCTTCTGACATCTTAAAGAT 360  
D 420 TATGCAACAGCTATTAATTTTGCMAAAAGGAAATTAACCTTCTGACATCTTAAAGAT 479  
Q 361 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGACTTCTACAGAGT 420  
D 480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGACTTCTACAGAGT 539  
Q 421 GAACCCGAAATTCCTTCTTGCAGAAACCACTGCTAGTGTCCAACTCTTAACTTTGA 480  
D 540 GAACCCGAAATTCCTTCTTGCAGAAACCACTGCTAGTGTCCAACTCTTAACTTTGA 599  
Q 481 ACTGTAGAACTCTGAGAGCAAGCAGCGATACAACTTAAAGAGCTGTGTACATT 540  
D 600 ACTGTAGAACTCTGAGAGCAAGCAGCGATACAACTTAAAGAGCTGTGTACATT 659  
Q 541 GAATGGAGTCTGATCTTCTGAGAGTACCGTTAATAGGCACTTATTCAGTGTGGA 600  
D 660 GAATGGAGTCTGATCTTCTGAGAGTACCGTTAATAGGCACTTATTCAGTGTGGA 719  
Q 601 GATCAAGAAATTTGTACAAATCACCCCTCAAGGAACGAGGATGAATCACTTTGGATTCT 660  
D 720 GATCAAGAAATTTGTACAAATCACCCCTCAAGGAACGAGGATGAATCACTTTGGATTCT 779  
Q 661 GCAAAAAAGGCTGCTGTGTAATTTTCTGAGAGCGATTAACAATTAAGTCAATCA 720  
D 780 GCAAAAAAGGCTGCTGTGTAATTTTCTGAGAGCGATTAACAATTAAGTCAATCA 839  
Q 721 CCCAGTAATTAATGATTTGAACACCACTGAGAAAGCTGACAGTGTGAGGCGATCCAGAAAG 780  
D 840 CCCAGTAATTAATGATTTGAACACCACTGAGAAAGCTGACAGTGTGAGGCGATCCAGAAAG 899  
Q 781 TATCAGAGGATGTTCTGTTTCAAACTTGATGTGAGCCATGTGGCACAAAATTAAGTCC 840  
D 900 TATCAGAGGATGTTCTGTTTCAAACTTGATGTGAGCCATGTGGCACAAAATTAAGTCC 959  
Q 841 ACCTATTCACAGATGAGAAACAGCACTTATTAATCACTTAAAGCAAGATGTAGAA 900  
D 960 ACCTATTCACAGATGAGAAACAGCACTTATTAATCACTTAAAGCAAGATGTAGAA 1019  
Q 901 AAGGCTGAATTTCTGTAAATAAAGCAAAACAGCTGCGTTAGCAAGAGCCAACTAATCAGA 960  
D 1020 AAGGCTGAATTTCTGTAAATAAAGCAAAACAGCTGCGTTAGCAAGAGCCAACTAATCAGA 1079  
Q 961 TGGGCTGGAATGAGGAAACATGTAAATGATAGGCGGAGCTCCAGCACAGAAAAAAGGTA 1020  
D 1080 TGGGCTGGAATGAGGAAACATGTAAATGATAGGCGGAGCTCCAGCACAGAAAAAAGGTA 1139  
Q 1021 GATCTGAATGCTGATCCCTGTGTGAGAAAAAAGATGGAATTAAGCAAGAACTGCCATGC 1080  
D 1140 GATCTGAATGCTGATCCCTGTGTGAGAAAAAAGATGGAATTAAGCAAGAACTGCCATGC 1199  
Q 1081 TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTAATATAGCAGATTTCAG 1140  
D 1200 TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTAATATAGCAGATTTCAG 1259  
Q 1141 AAAGTTAATGAGTGTCTTTCAGAAAGTGAATCTGTAGGTTCTGATGACTCAATGAT 1200  
D 1260 AAAGTTAATGAGTGTCTTTCAGAAAGTGAATCTGTAGGTTCTGATGACTCAATGAT 1319  
Q 1201 GGGGAGTCTGAATCAAAATGCCAAAGTGAATGATGTATTTGAGAGTCTTAAATGAGTAGAT 1260  
D 1320 GGGGAGTCTGAATCAAAATGCCAAAGTGAATGATGTATTTGAGAGTCTTAAATGAGTAGAT 1379

Q 1261 GAATATTTCTGTTCTTACAGAAATAGACTTACGGCCAGTATCCATGAGGCTTTA 1320  
D 1380 GAATATTTCTGTTCTTACAGAAATAGACTTACGGCCAGTATCCATGAGGCTTTA 1439  
Q 1321 ATATGTAAGAAGTGAAGAGTTCACCTCAATCAGTAGAGAGTAAATTTGAAGACAAATA 1380  
D 1440 ATATGTAAGAAGTGAAGAGTTCACCTCAATCAGTAGAGAGTAAATTTGAAGACAAATA 1499  
Q 1381 TTTGGGAAAACCTATGGAAGAAAGCAAGCTTCCCACTTAAGCATCTAAGTAAAT 1440  
D 1500 TTTGGGAAAACCTATGGAAGAAAGCAAGCTTCCCACTTAAGCATCTAAGTAAAT 1559  
Q 1441 CTAAATATAGAGCATTTGTTTACTGAGCCACAGATATATCAAGAGGTCCTTCAAAAT 1500  
D 1560 CTAAATATAGAGCATTTGTTTACTGAGCCACAGATATATCAAGAGGTCCTTCAAAAT 1619  
Q 1501 AAATTAAGGCTTAAAGAGACCTACATGAGGCTTCACTCTGAGGATTTTATCAAGAA 1560  
D 1620 AAATTAAGGCTTAAAGAGACCTACATGAGGCTTCACTCTGAGGATTTTATCAAGAA 1679  
Q 1561 GCAGATTTGGCAGTTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTACCAGGAG 1620  
D 1680 GCAGATTTGGCAGTTTCAAAAAGACTCCTGAATGATTAATCAGGAACTTACCAGGAG 1739  
Q 1621 CAGATGTGTCAGTGAATGATTAATTAATAGTGTGATGAGATTAACCAAGAGTAT 1680  
D 1740 CAGATGTGTCAGTGAATGATTAATTAATAGTGTGATGAGATTAACCAAGAGTAT 1799  
Q 1681 TCTATTCAGATGAGAAAAATCCTTAACCAATAGATCACTGAAAAAGAAATCTCTTC 1740  
D 1800 TCTATTCAGATGAGAAAAATCCTTAACCAATAGATCACTGAAAAAGAAATCTCTTC 1859  
Q 1741 AAACGAAAGCTAAGCTTAATAGCAGACTATTAAGCAATATGTGAATCTGAATTAATATC 1800  
D 1860 AAACGAAAGCTAAGCTTAATAGCAGACTATTAAGCAATATGTGAATCTGAATTAATATC 1919  
Q 1801 CACAATTCAAAAGCAGCTTAAAAAGATGAGTGTGAGAGGAGTCTTACAGGAGTAT 1860  
D 1920 CACAATTCAAAAGCAGCTTAAAAAGATGAGTGTGAGAGGAGTCTTACAGGAGTAT 1979  
Q 1861 CATGCGCTTGAACCTAGTACGATGAGTAAAGTAAAGCACTGCAACTGAGAGTAAAG 1920  
D 1980 CATGCGCTTGAACCTAGTACGATGAGTAAAGTAAAGCACTGCAACTGAGAGTAAAG 2039  
Q 1921 ATTGATAGTGTCTGAGAGTGAAGATTAAGGAAAAAAGTAAACCAATGCGCAATC 1980  
D 2040 ATTGATAGTGTCTGAGAGTGAAGATTAAGGAAAAAAGTAAACCAATGCGCAATC 2099  
Q 1981 AGGACAGAGAAACCTACAGATGAGTAAAGTAAAGCACTGCAACTGAGAGTAAAG 2040  
D 2100 AGGACAGAGAAACCTACAGATGAGTAAAGTAAAGCACTGCAACTGAGAGTAAAG 2159  
Q 2041 AGTAAACAGCCAAATGAACAGCAAGTAAAGCACTGAGAGTAAAGTAAAGTAAAG 2100  
D 2160 AGTAAACAGCCAAATGAACAGCAAGTAAAGCACTGAGAGTAAAGTAAAGTAAAG 2219  
Q 2101 AAGTTAACAATGCAACCTGCTCTTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2160  
D 2220 AAGTTAACAATGCAACCTGCTCTTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2279  
Q 2161 TTTGTCAATCTAGCTTCCAGAGAGAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2220  
D 2280 TTTGTCAATCTAGCTTCCAGAGAGAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2339  
Q 2221 TCTAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAACT 2280  
D 2340 TCTAATATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAACT 2399  
Q 2281 GAAAGATCTGTAGAGATGAGTATTTCAATTTGATGCTGATGATTAAGGACATCAG 2340  
D 2400 GAAAGATCTGTAGAGATGAGTATTTCAATTTGATGCTGATGATTAAGGACATCAG 2459

2341 GAAATATCTCGTTACTGGAAGTTAGCACTAGGGAGGCAAAAACAGAACCAATATAA 2400  
2460 GAAAGATATCTCGTTACTGGAAGTTAGCACTAGGGAGGCAAAAACAGAACCAATATAA 2519  
2401 TGTGTAGTCAGTGTGACAGATTTGAAACCCCAAGGACTAATTCATGTTTCCAAA 2460  
2520 TGTGTAGTCAGTGTGACAGATTTGAAACCCCAAGGACTAATTCATGTTTCCAAA 2579  
2461 GATAATAGAAATGACACAGAGGCTTTAAGTATCCATTGGACATGAAAGTTAACACAGT 2520  
2580 GATAATAGAAATGACACAGAGGCTTTAAGTATCCATTGGACATGAAAGTTAACACAGT 2539  
2521 CGGGAACACAGCATAGAAATGAGAAAGTGAACCTGATGCTAGTATTTGCAAAATACA 2580  
2640 CGGGAACACAGCATAGAAATGAGAAAGTGAACCTGATGCTAGTATTTGCAAAATACA 2699  
2581 TTCAGGTTTCAAGGCGCAGATTTGCTCTGTTTTCAAATCCAGAAATGCAAGAG 2640  
2700 TTCAGGTTTCAAGGCGCAGATTTGCTCTGTTTTCAAATCCAGAAATGCAAGAG 2759  
2641 GAATGTGCAACATTTCTGCCCACCTGGGCTCTTAAGAAACAAAGTCCAAAAGTCACT 2700  
2760 GATGTGCAACATTTCTGCCCACCTGGGCTCTTAAGAAACAAAGTCCAAAAGTCACT 2819  
2701 TTTGAATGTGAACAAAAGAGAAATCAAGAAAGATGAGTCTAATATCAAGCCTGTA 2760  
2820 TTTGAATGTGAACAAAAGAGAAATCAAGAAAGATGAGTCTAATATCAAGCCTGTA 2879  
2761 CAGACGTAATTTCACTGACGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2820  
2880 CAGACGTAATTTCACTGACGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2939  
2821 AATGCCAAATGTAGTATCAAGAGGCTAGGTTTGTCTATCATCTCAGTCCAGAGG 2880  
2940 AATGCCAAATGTAGTATCAAGAGGCTAGGTTTGTCTATCATCTCAGTCCAGAGG 2999  
2881 AACGAAATGTAGTATCACTGCAAAATCAATGACATTTTACAAAACCCATATCTGATA 2940  
3000 AACGAAATGTAGTATCACTGCAAAATCAATGACATTTTACAAAACCCATATCTGATA 3059  
2941 CCACCACTTTTCCATCAAGTATTTGTTAAACATAAATGTAAAGAAATCTGCTAGAG 3000  
3060 CCACCACTTTTCCATCAAGTATTTGTTAAACATAAATGTAAAGAAATCTGCTAGAG 3119  
3001 GAAACCTTTGAGAACATTTCAATGTCACCTGAAAGAGAAATGGAATGGAATGCAATTC 3060  
3120 GAAACCTTTGAGAACATTTCAATGTCACCTGAAAGAGAAATGGAATGGAATGCAATTC 3179  
3061 AGTACGTAGACCAATTTAGCCGTAAATTAACATTAGAGAAATGTTTAAAGAGCCAGC 3120  
3180 AGTACGTAGACCAATTTAGCCGTAAATTAACATTAGAGAAATGTTTAAAGAGCCAGC 3239  
3121 TCAGCAATTAATTAAGTAGGTTCCAGTAACTAATGAAGTGGCCGATATTAATGA 3180  
3240 TCAGCAATTAATTAAGTAGGTTCCAGTAACTAATGAAGTGGCCGATATTAATGA 3299  
3181 AATAGTTCAGTATGAAAAAATTCAGAGCAAGTAGTAGAAACAGAGGCCAAATTTG 3240  
3300 AATAGTTCAGTATGAAAAAATTCAGAGCAAGTAGTAGAAACAGAGGCCAAATTTG 3359  
3241 AATGCTATGCTTAGATTTAGGGGTTTTCGAACCTGAGGCTATTAACAAAGTCTCTCGA 3300  
3360 AATGCTATGCTTAGATTTAGGGGTTTTCGAACCTGAGGCTATTAACAAAGTCTCTCGA 3419  
3301 AGTAATGTAGATCTGAAATTAATAAAGCAAGATTAAGAAATGATTTAGACTGTT 3360  
3420 AGTAATGTAGATCTGAAATTAATAAAGCAAGATTAAGAAATGATTTAGACTGTT 3479  
3361 AATACAGATTTCTCCATATCTGATTTAGATTAACATTAAGACAGCTATGGAAATAGT 3420  
3480 AATACAGATTTCTCCATATCTGATTTAGATTAACATTAAGACAGCTATGGAAATAGT 3539  
3421 CATGATCTCAGGTTTCTGAGACACCTGATGACCTGTTAGATGAGTGAATTAAG 3480

3540 CATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGAATTAAG 3599  
3481 GAAGATATCTAGTTTGTCTGAAATGACATTAAGAAAGTCTGCTGTTTACGAAAGC 3540  
3600 GAAGATATCTAGTTTGTCTGAAATGACATTAAGAAAGTCTGCTGTTTACGAAAGC 3659  
3541 GTCCAGAAAGAGAGCTTACAGAGTCTAGGCCCTTTCACCATACATTTGGGCTCAG 3600  
3660 GTCCAGAAAGAGAGCTTACAGAGTCTAGGCCCTTTCACCATACATTTGGGCTCAG 3719  
3601 GGTTCACGAAAGAGGCGCAAGAAATTAAGTCTCAGAAAGAACTTATCTAGAGAT 3660  
3720 GGTTCACGAAAGAGGCGCAAGAAATTAAGTCTCAGAAAGAACTTATCTAGAGAT 3779  
3661 GAAGAGCTTCCCTGCTTCCACACCTGTTATTTGGTAAGTAAACATATACCTTCTCAG 3720  
3780 GAAGAGCTTCCCTGCTTCCACACCTGTTATTTGGTAAGTAAACATATACCTTCTCAG 3839  
3721 TCTACTAGCATAGCACCCGTTGCTACGAGTGTCTGTCTAAGAACACAGAGAAATTA 3780  
3840 TCTACTAGCATAGCACCCGTTGCTACGAGTGTCTGTCTAAGAACACAGAGAAATTA 3899  
3781 TTATCATTTGAAGATAGCTTAAATGACTGCAATACAGGTAATTTGGCAAGCATCT 3840  
3900 TTATCATTTGAAGATAGCTTAAATGACTGCAATACAGGTAATTTGGCAAGCATCT 3959  
3841 CAGGAATCATCTTGTAGTGAAGAAACAAATGTTGCTGATGCTGTTTCTTACAGTGC 3900  
3960 CAGGAATCATCTTGTAGTGAAGAAACAAATGTTGCTGATGCTGTTTCTTACAGTGC 4019  
3901 AGTGAATTTGAAGATAGCTTAAATGACTGCAATACAGGTAATTTGGCAAGCATCT 3960  
4020 AGTGAATTTGAAGATAGCTTAAATGACTGCAATACAGGTAATTTGGCAAGCATCT 4079  
3961 TCCAAACAAATGAGGATGATCTGAAAGCCAGAGGTTGCTGATGATCAAGAAATG 4020  
4080 TCCAAACAAATGAGGATGATCTGAAAGCCAGAGGTTGCTGATGATCAAGAAATG 4139  
4021 GTTTCAGATGATGAAGAAAGAGAGAGGCTTGAAGAAATTAATCAAGAGAGCAAGC 4080  
4140 GTTTCAGATGATGAAGAAAGAGAGGCTTGAAGAAATTAATCAAGAGAGCAAGC 4199  
4081 ATGGAATTTCAAACTTAGGTGAAGAGCATCTGGGTGTGAGAGTGAAGAAACGCTCTGAA 4140  
4200 ATGGAATTTCAAACTTAGGTGAAGAGCATCTGGGTGTGAGAGTGAAGAAACGCTCTGAA 4259  
4141 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACAATCAGCAGAGAGGATACCATG 4200  
4260 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACAATCAGCAGAGAGGATACCATG 4319  
4201 CAACATTAACCTGATAAGGCTCCAGAGGAAATGGCTGAAGTCAACCTGTTAGAACG 4260  
4320 CAACATTAACCTGATAAGGCTCCAGAGGAAATGGCTGAAGTCAACCTGTTAGAACG 4379  
4261 CATGGAGGAGGCTTCTTAACAGCTACCTTCCATATAAGTACTCTTGCCTTGAAG 4320  
4380 CATGGAGGAGGCTTCTTAACAGCTACCTTCCATATAAGTACTCTTGCCTTGAAG 4439  
4321 GACTGCGAAATTCAGAACAGACATCAGAAAAGAGATTTAACTTACAGAAAAGT 4380  
4440 GACTGCGAAATTCAGAACAGACATCAGAAAAGAGATTTAACTTACAGAAAAGT 4499  
4381 AGTGAATTAACCTTAAGCCAGAAATCCAGAAAGGCTTTCGCTGACAAAGTTGAGTGTCT 4440  
4500 AGTGAATTAACCTTAAGCCAGAAATCCAGAAAGGCTTTCGCTGACAAAGTTGAGTGTCT 4559  
4441 GCAGATAGTTTCAACAGTAAAAATTAAGAACAGAGAGTGAAGAGTCAATCCCTTTAA 4500  
4560 GCAGATAGTTTCAACAGTAAAAATTAAGAACAGAGAGTGAAGAGTCAATCCCTTTAA 4619  
4501 TGCCATCAATTAAGATAGTGGTGTACATGACAGTTGCTCTGAGAGTCTTCAAGATTAAG 4560



```

Db 4620 TGCCCATCATAGATAGATAGTGTACATGCACAGTTGCTTGGAGTCTTCAGAAATAGA 4679
Oy 4561 AACTACCCATCTCAAGAGAGCTCATTAAGTGTGTGATGTGAGAGCAACAGCTGGAA 4620
Db 4680 AACTACCCATCTCAAGAGAGCTCATTAAGTGTGTGATGTGAGAGCAACAGCTGGAA 4739
Oy 4621 GAGTCTGGGCGACACAGATTTGACGAAACATCTTACTTGGCAAGGCAAGATAGAGGA 4680
Db 4740 GAGTCTGGGCGACACAGATTTGACGAAACATCTTACTTGGCAAGGCAAGATAGAGGA 4799
Oy 4681 ACCCTTACTTGAATCTGGAATCAGCTCTTCTGTGATGACCTGATATCTCTCT 4740
Db 4800 ACCCTTACTTGAATCTGGAATCAGCTCTTCTGTGATGACCTGATATCTCTCT 4859
Oy 4741 GAAGACAGAGCCCGACAGTCTGCTGTGGCAATATCCATCTTCAACCTGTCATG 4800
Db 4860 GAAGACAGAGCCCGACAGTCTGCTGTGGCAATATCCATCTTCAACCTGTCATG 4919
Oy 4801 AAAGTTCCCAATTGAAAGTTCGAGAACTGCCAGAGTCCAGCTGCTGCTCATACT 4860
Db 4920 AAAGTTCCCAATTGAAAGTTCGAGAACTGCCAGAGTCCAGCTGCTGCTCATACT 4979
Oy 4861 GATACTGCTGGGTATATGATCAATGGAAGAGTGTAGCAGGAGAGCAACGATTTGACA 4920
Db 4980 GATACTGCTGGGTATATGATCAATGGAAGAGTGTAGCAGGAGAGCAACGATTTGACA 5039
Oy 4921 GCTTCAACAGAAAGGCTCAACAAAGATGCTGATGCTGCTGCTGACCCAGAA 4980
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGATGCTGATGCTGCTGCTGACCCAGAA 5099
Oy 4981 GAATTTATGCTGCTGATCAAGTTTGCAGAAAAACACACATCATCTTAACTAATCTAAT 5040
Db 5100 GAATTTATGCTGCTGATCAAGTTTGCAGAAAAACACACATCATCTTAACTAATCTAAT 5159
Oy 5041 ACTGAAGAGACATCATCTGTTGTAAGAAACAGATGCTGATGCTGCTGACAGGACA 5100
Db 5160 ACTGAAGAGACATCATCTGTTGTAAGAAACAGATGCTGATGCTGCTGACAGGACA 5219
Oy 5101 CTGAATATATTTCTAGGAATTCGAGAGGAAAAATGGTATGATTAATTTCTGGTGACC 5160
Db 5220 CTGAATATATTTCTAGGAATTCGAGAGGAAAAATGGTATGATTAATTTCTGGTGACC 5279
Oy 5161 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGAGCATGATTTGAAGTCAAGAGAGATGTG 5220
Db 5280 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGAGCATGATTTGAAGTCAAGAGAGATGTG 5339
Oy 5221 GTCATGGAAGAAACCAACCAAGTCCAAAGCGACGAAAGAAATCCGAGACAGAAAGATC 5280
Db 5340 GTCATGGAAGAAACCAACCAAGTCCAAAGCGACGAAAGAAATCCGAGACAGAAAGATC 5399
Oy 5281 TTCAGGGGGCTGGAATCTGTTGCTATGAGGCGCTTCCACCAACATGCCACAGATCAACTG 5340
Db 5400 TTCAGGGGGCTGGAATCTGTTGCTATGAGGCGCTTCCACCAACATGCCACAGATCAACTG 5459
Oy 5341 GAATGATGGAAGAGAGTGTGCTGATGAGGCGCTTCCACCAACATGCCACAGATCAACTG 5400
Db 5460 GAATGATGGAAGAGAGTGTGCTGATGAGGCGCTTCCACCAACATGCCACAGATCAACTG 5519
Oy 5401 GGCACAGGAGTCCACCAATTTGGTGTGACGACGAGATGCCGCGACAGAGACATGGC 5460
Db 5520 GGCACAGGAGTCCACCAATTTGGTGTGACGACGAGATGCCGCGACAGAGACATGGC 5579
Oy 5461 TTCATGCAATTTGGGAGAGATGTGTAGGACCTGTGTGACCCGAGAGTGGTGTGGAC 5520
Db 5580 TTCATGCAATTTGGGAGAGATGTGTAGGACCTGTGTGACCCGAGAGTGGTGTGGAC 5639
Oy 5521 AGTGTAGCACTGTACCAAGTGTGAGAGAGTGTGACACCTGATATACCCAGATCCCCCAG 5580
Db 5640 AGTGTAGCACTGTACCAAGTGTGAGAGAGTGTGACACCTGATATACCCAGATCCCCCAG 5699
Oy 5581 AGCCACTAC 5589
Db 5700 AGCCACTAC 5708

```

```

RESULT 13
AA287995
ID AA287995 standard; DNA: 5712 BP.
XX
AC AA287995:
XX
DT 06-JUN-2000 (first entry)
XX
DE BRCA1 gene sequence (GenBank Accn No: U14680).
XX
KW BRCA1; tumour; prostate cancer; cytostatic; antiproliferative;
KM gene therapy; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT /tag= a
FT /transl_except= (pos: 1452..1454, aa:asp)
FT /transl_except= (pos: 1482..1484, aa:asp)
FT /transl_except= (pos: 1581..1583, aa:ser)
FT /transl_except= (pos: 1899..1901, aa:glu)
FT /transl_except= (pos: 1902..1904, aa:leu)
FT /transl_except= (pos: 1905..1907, aa:glu)
FT /transl_except= (pos: 1908..1910, aa:leu)
FT /transl_except= (pos: 1911..1913, aa:asn)
FT /transl_except= (pos: 1914..1916, aa:ile)
FT /transl_except= (pos: 1917..1919, aa:met)
FT /transl_except= (pos: 4476..4478, aa:val)
FT /transl_except= (pos: 4479..4481, aa:leu)
FT /transl_except= (pos: 4482..4484, aa:glu)
FT /transl_except= (pos: 4533..4535, aa:tyr)
FT /transl_except= (pos: 4689..4691, aa:pro)
FT /note= "Xaa= unknown"
XX
CA2218197-A1.
XX
PD 05-JUN-1999.
XX
PE 12-DEC-1997; 97CA-2218197.
XX
PR 05-DEC-1997; 97US-0986106.
XX
PA (UNIW ) UNIV WASHINGTON.
PI (UYVA-) UNIV VANDERBILT.
PI Robinson-Benion CL, Thompson ME, Holt JT, Jensen RA, Steiner MS;
PI King M;
XX
DR WPI: 2000-238071/21.
XX
P-PSDB; AAT77818.
XX
PT New method of treatment and suppression of prostate cancer comprises
PT using the BRCA family of genes to decrease the growth rate of the tumor
PT
XX
PS Claim 11: Page 98-104; 166pp; English.
XX
CC The invention relates to a method for suppressing the growth of a
CC prostate tumour in a mammal that comprises introducing to the tumour a
CC vector comprising a nucleic acid sequence encoding a BRCA family gene
CC product operatively linked to a promoter, where production of the BRCA
CC family gene product results in a decrease in the growth rate of the
CC tumour. The methods are used to suppress the growth of and also to treat
CC prostate tumour in a mammal where the tumour is gene-linked hereditary
CC prostate cancer or sporadic prostate cancer. The present sequence
CC represents a BRCA1 gene sequence (GenBank Accn No: U14680) that can be
CC administered for suppressing tumour.
XX
SQ Sequence 5712 BP; 1956 A; 1099 C; 1274 G; 1383 T; 0 other;

```



Query Match 99.9%; Score 5585.8; DB 21; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 60  
DB 120 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 179  
QY 61 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 120  
DB 180 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 239  
QY 121 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 180  
DB 240 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 239  
QY 181 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 240  
DB 300 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 359  
QY 241 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 300  
DB 360 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 419  
QY 301 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 360  
DB 420 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 479  
QY 361 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 420  
DB 480 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 539  
QY 421 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 480  
DB 540 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 539  
QY 481 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 540  
DB 600 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 659  
QY 541 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 600  
DB 660 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 719  
QY 601 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 660  
DB 720 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 779  
QY 661 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 720  
DB 780 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 839  
QY 721 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 780  
DB 840 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 899  
QY 781 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 840  
DB 900 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 959  
QY 841 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 900  
DB 960 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1019  
QY 901 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 960  
DB 1020 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1079  
QY 961 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1020  
DB 1080 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1139  
QY 1021 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1080

DB 1140 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1199  
QY 1081 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1140  
DB 1200 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1259  
QY 1141 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1200  
DB 1260 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1319  
QY 1201 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1260  
DB 1320 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1379  
QY 1261 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1320  
DB 1380 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1439  
QY 1321 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1380  
DB 1440 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1499  
QY 1381 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1440  
DB 1500 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1559  
QY 1441 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1500  
DB 1560 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1619  
QY 1501 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1560  
DB 1620 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1679  
QY 1561 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1620  
DB 1680 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1739  
QY 1621 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1680  
DB 1740 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1799  
QY 1681 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1740  
DB 1800 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1859  
QY 1741 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1800  
DB 1860 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1919  
QY 1801 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1860  
DB 1920 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1979  
QY 1861 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1920  
DB 1980 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2039  
QY 1921 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 1980  
DB 2040 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2099  
QY 1981 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2040  
DB 2100 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2159  
QY 2041 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2100  
DB 2160 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2219  
QY 2101 AAGGATTTATCTGCTTGGCGTTGAAGAAGTACCAAAATGCTATTATGCTATGCGAAA 2160

```
Db 2220 AAGTTAACAAATGCACCTGGTCTTTTACTAAGTGTCAATACACAGTGAACCTTAAAGAA 2279
QY 2161 TTGTCAATCCAGCTTCCAAAGAGAGAAAAAAGAGAAATGAAACAGTTAAAGTG 2220
Db 2280 TTTGTCAATCCTAGCTTCCAAAGAGAGAAAAAAGAGAAATGAAACAGTTAAAGTG 2339
QY 2221 TCTAATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGGCAACT 2280
Db 2340 TCTAATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGGCAACT 2399
QY 2281 GAAAGATCTGTAGAGAGTACAGATTTTCATTTGGTACCTGGTACTGATTAATGGCACTCG 2340
Db 2400 GAAAGATCTGTAGAGAGTACAGATTTTCATTTGGTACCTGGTACTGATTAATGGCACTCG 2459
QY 2341 GAAAGTATCTGCTTCTGGAAGTACAGCTTAGGAGAGGCAAAAAACAGAACCAATAA 2400
Db 2460 GAAAGTATCTGCTTCTGGAAGTACAGCTTAGGAGAGGCAAAAAACAGAACCAATAA 2519
QY 2401 TGTGTAGTCAAGTGTGACAGCATTTTGAACCCCAAGGACATAATTCATGGTTGTTCCAAA 2460
Db 2520 TGTGTAGTCAAGTGTGACAGCATTTTGAACCCCAAGGACATAATTCATGGTTGTTCCAAA 2579
QY 2461 GATATATATAAATGACACAGAGAGGCTTTAAGTATCCATTGGGACATGAACTTAACCACT 2520
Db 2580 GATATATATAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAACTTAACCACT 2639
QY 2521 CGGGAACACAGCATGGAATGGAAGAGAGTGAACCTTGATGCTCAGTATTTGAGAAATACA 2580
Db 2640 CGGGAACACAGCATGGAATGGAAGAGAGTGAACCTTGATGCTCAGTATTTGAGAAATACA 2699
QY 2581 TTCAAGGTTTCAAGCGCCAGTCAATTTGCTGTCTTTCAATCCAGGAATCAGAAAG 2640
Db 2700 TTCAAGGTTTCAAGCGCCAGTCAATTTGCTGTCTTTCAATCCAGGAATCAGAAAG 2759
QY 2641 GAATGTGCAATCTCTGCCACTCTGGCTCTGCTTAAAGAAACAAAGTCCAAAGCACT 2700
Db 2760 GAATGTGCAATCTCTGCCACTCTGGCTCTGCTTAAAGAAACAAAGTCCAAAGCACT 2819
QY 2701 TTTGATGTGAACAAAGAGAGAAATATCAAGGAAAGATGACTTAATATCAAGCCTGTA 2760
Db 2820 TTTGATGTGAACAAAGAGAGAAATATCAAGGAAAGATGACTTAATATCAAGCCTGTA 2879
QY 2761 CAGACAGTTAATATCTACTGCAAGGCTTCTGTGTGGTGGTCAGAAAGATPAGCCAGTTGAT 2820
Db 2880 CAGACAGTTAATATCTACTGCAAGGCTTCTGTGTGGTGGTCAGAAAGATPAGCCAGTTGAT 2939
QY 2821 AATGCAATGATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCCAGAGC 2880
Db 2940 AATGCAATGATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCCAGAGC 2999
QY 2881 AACGAACTGAGCTACTTACTCCAAATAAACATGACCTTTTACAAACCCCATATGCTATA 2940
Db 3000 AACGAACTGAGCTACTTACTCCAAATAAACATGACCTTTTACAAACCCCATATGCTATA 3059
QY 2941 CCACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTPAGAAAAATCTGCTAGAG 3000
Db 3060 CCACCACTTTTCCCATCAAGTCAATTTGTTAAACTAAATGTPAGAAAAATCTGCTAGAG 3119
QY 3001 GAAAACTTTGAAGAACATCAATGTCACCTGAAAGAGAAATGGGAAATGAGAAACNTCCA 3060
Db 3120 GAAAACTTTGAAGAACATCAATGTCACCTGAAAGAGAAATGGGAAATGAGAAACNTCCA 3179
QY 3061 AGTACAGTACAGACATTAAGCGGTATTAACATTTAGAGAAATGTTTAAAGAGCCAC 3120
Db 3180 AGTACAGTACAGACATTAAGCGGTATTAACATTTAGAGAAATGTTTAAAGAGCCAC 3239
QY 3121 TCAAGCAATATTAATGAAGTAGTTCAGTACTAAATGAAGTGGGCTCCAGTATTAATGA 3180
Db 3240 TCAAGCAATATTAATGAAGTAGTTCAGTACTAAATGAAGTGGGCTCCAGTATTAATGA 3299
QY 3181 ATAGGTTCCAGTGAAGAAACATTTACAGCAGAACTAGTGAAGAAACAGAGGCGCAAAATG 3240
Db 3300 ATAGGTTCCAGTGAAGAAACATTTACAGCAGAACTAGTGAAGAAACAGAGGCGCAAAATG 3359

QY 3241 AATGCTATGCTAGATTAAGGGGTTTGGCAACCTGAGTCTATAAACAAGCTTCTCTGA 3300
Db 3360 AATGCTATGCTAGATTAAGGGGTTTGGCAACCTGAGTCTATAAACAAGCTTCTCTGA 3419
QY 3301 ACTAATTTAGCATCTGAAATTAAGAAACAGAAATATGAAGAGTATGTTACAGCTGT 3360
Db 3420 ACTAATTTAGCATCTGAAATTAAGAAACAGAAATATGAAGAGTATGTTACAGCTGT 3479
QY 3361 AATACAGATTTCTCCAAATGATTTAGTAACTTGAACAGCCATAGGAGTAGT 3420
Db 3480 AATACAGATTTCTCCAAATGATTTAGTAACTTGAACAGCCATAGGAGTAGT 3539
QY 3421 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGAAATTAAG 3480
Db 3540 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGAAATTAAG 3599
QY 3481 GAAGATCTAGTTTGTCTGAAATATGACATTAAGAAAGTCTGCTGTTTACGAAAGC 3540
Db 3600 GAAGATCTAGTTTGTCTGAAATATGACATTAAGAAAGTCTGCTGTTTACGAAAGC 3659
QY 3541 GTCCAGAAAGAGAGCTTAGCAGAGTCTGACCTGACCTTACCCATACACATTTGGCTCAG 3600
Db 3660 GTCCAGAAAGAGAGCTTAGCAGAGTCTGACCTGACCTTACCCATACACATTTGGCTCAG 3719
QY 3601 GGTACCCGAGAGGCGCCAGAAATTAGAGTCTCAGAGAGAACTTATCTAGAGAT 3660
Db 3720 GGTACCCGAGAGGCGCCAGAAATTAGAGTCTCAGAGAGAACTTATCTAGAGAT 3779
QY 3661 GAAGAGCTTCCCTGCTCCAAACACTTGTATTTGGTAAAGTAAACATATACCTTCCAG 3720
Db 3780 GAAGAGCTTCCCTGCTCCAAACACTTGTATTTGGTAAAGTAAACATATACCTTCCAG 3839
QY 3721 TCTACTAGGCAATAGACCGTTGCTACGAGTGTCTGTAAAGAACACAGAGAGAAATTTA 3780
Db 3840 TCTACTAGGCAATAGACCGTTGCTACGAGTGTCTGTAAAGAACACAGAGAGAAATTTA 3899
QY 3781 TTATCATTTGAAGAAATAGCTTAAATGATCTGAGTAACCAAGTAAATTTGGCAAGGATCT 3840
Db 3900 TTATCATTTGAAGAAATAGCTTAAATGATCTGAGTAACCAAGTAAATTTGGCAAGGATCT 3959
QY 3841 CAGGAACATCACTTACTAGTGAAGAAACAAATTTCTGCTACTGTTTCTTCCACAGTGC 3900
Db 3960 CAGGAACATCACTTACTAGTGAAGAAACAAATTTCTGCTACTGTTTCTTCCACAGTGC 4019
QY 3901 AGTGAATTTGAAGACTTACTGCAATATCAAAACACCAGATCTCTTCTTATGATGTTCT 3960
Db 4020 AGTGAATTTGAAGACTTACTGCAATATCAAAACACCAGATCTCTTCTTATGATGTTCT 4079
QY 3961 TCCAAACAAATGAGCATCACTGTAAGCCAGGAGTTGGTCTGAGTACAGAGAAATG 4020
Db 4080 TCCAAACAAATGAGCATCACTGTAAGCCAGGAGTTGGTCTGAGTACAGAGAAATG 4139
QY 4021 GTTTCAGATGATGAAGAAAGAGAGAGCGGCTTGGAGAAAGAAATATCAAGAGAGCAAGC 4080
Db 4140 GTTTCAGATGATGAAGAAAGAGAGAGCGGCTTGGAGAAAGAAATATCAAGAGAGCAAGC 4199
QY 4081 AATGATTTCAAACTTGTGTGAAGCAGCATCTGGGTGTGAGATGGAACAGCGCTCTGAA 4140
Db 4200 AATGATTTCAAACTTGTGTGAAGCAGCATCTGGGTGTGAGATGGAACAGCGCTCTGAA 4259
QY 4141 GACTGCTGAGGCTATCTCTCAGAGTACATTTTAAACACTCAGCAGAGAGATACCAT 4200
Db 4260 GACTGCTGAGGCTATCTCTCAGAGTACATTTTAAACACTCAGCAGAGAGATACCAT 4319
QY 4201 CAACATTAACCTGATTAAGCTCCAGCAGGAATGCTGTAAGTGAAGCTGTGTTAAGACAG 4260
Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGGAATGCTGTAAGTGAAGCTGTGTTAAGACAG 4379
QY 4261 CATGGAGCGAGCCTTCTTAACAGTACCTCCATCAATATGACTCTTGCCCTTGAG 4320
Db 4380 CATGGAGCGAGCCTTCTTAACAGTACCTCCATCAATATGACTCTTGCCCTTGAG 4439
```

```

QY 4321 GACCTGGGAATTCAGAACAAAGCAGATCAGAAAAAGCACTATTAACTTCAGAGAAAGT 4380
    |||
Db 4440 GACCTGGGAATTCAGAACAAAGCAGATCAGAAAAAGCACTATTAACTTCAGAGAAAGT 4499
QY 4381 AGTGAATACCCCTTAAGCCAGAAATCCAGAAAGGCTTTCTCTGACAAAGTTTGAGGTCT 4440
    |||
Db 4500 AGTGAATACCCCTTAAGCCAGAAATCCAGAAAGGCTTTCTCTGACAAAGTTTGAGGTCTCT 4559
QY 4441 GCAGATAGTTCTACAGATAAATAAAGAACAGAGAGTGGAAAGGTATCCCTCTTAAA 4500
    |||
Db 4560 GCAGATAGTTCTACAGATAAATAAAGAACAGAGAGTGGAAAGGTATCCCTCTTAAA 4619
QY 4501 TGGCCATCATTTAGATGATAGTGGTACATCAGAGTTGCTCTGGAGCTTTCAGAAATGA 4560
    |||
Db 4620 TGGCCATCATTTAGATGATAGTGGTACATCAGAGTTGCTCTGGAGCTTTCAGAAATGA 4679
QY 4561 AACTACCATCTCAAGAGAGCTCATTAAGTTGTTATGTGGAGAGCAACAGCTGGAA 4620
    |||
Db 4680 AACTACCATCTCAAGAGAGCTCATTAAGTTGTTATGTGGAGAGCAACAGCTGGAA 4739
QY 4621 GAGTCTGGCCACACAGATTTGACGAAACATCTTACTTGGCAAGCAGATAGAGGA 4680
    |||
Db 4740 GAGTCTGGCCACACAGATTTGACGAAACATCTTACTTGGCAAGCAGATAGAGGA 4799
QY 4681 ACCCTTACCTGGAATCTGGAATCAGCCTCTCTCTGATGACCCCTGAATCTGATCTTCT 4740
    |||
Db 4800 ACCCTTACCTGGAATCTGGAATCAGCCTCTCTCTGATGACCCCTGAATCTGATCTTCT 4859
QY 4741 GAAGACAGAGCCCAAGAGTACGCTCGTTGGCAACATACCATTTTACCTTCGATG 4800
    |||
Db 4860 GAAGACAGAGCCCAAGAGTACGCTCGTTGGCAACATACCATTTTACCTTCGATG 4919
QY 4801 AAAGTTCCTCCCAATTTGAAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCATCTACT 4860
    |||
Db 4920 AAAGTTCCTCCCAATTTGAAAGTTGCAAGATCTGCCAGAGTCCAGCTGCTCATCTACT 4979
QY 4861 GATACGTCTGGTATATATGCAATGGAAGAAAGTGTGAGGAGGAGAAAGCAGAAATTGACA 4920
    |||
Db 4980 GATACGTCTGGTATATATGCAATGGAAGAAAGTGTGAGGAGGAGAAAGCAGAAATTGACA 5039
QY 4921 GCTTCAACAGAAAGGGTCAACAAAGAAATCTCATGTGTGCTGCTGACCCAGAA 4980
    |||
Db 5040 GCTTCAACAGAAAGGGTCAACAAAGAAATCTCATGTGTGCTGCTGACCCAGAA 5099
QY 4981 GAATTTATGCTGCTGATAGAGTTTGGCAGAAAAACACACATCACTTAATAATCTAATT 5040
    |||
Db 5100 GAATTTATGCTGCTGATAGAGTTTGGCAGAAAAACACACATCACTTAATAATCTAATT 5159
QY 5041 ACTGAAGAGACTACTCATGTGTTATGAAAGACAGATGCTGAGTTTGTGTGAACGAGACA 5100
    |||
Db 5160 ACTGAAGAGACTACTCATGTGTTATGAAAGACAGATGCTGAGTTTGTGTGAACGAGACA 5219
QY 5101 CTGAATATTTTATAGAAATTTGGGAGAGAAATGGGTAGTTGCTATTTCTGGGTGACC 5160
    |||
Db 5220 CTGAATATTTTATAGAAATTTGGGAGAGAAATGGGTAGTTGCTATTTCTGGGTGACC 5279
QY 5161 CACTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATG 5220
    |||
Db 5280 CACTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATG 5339
QY 5221 GTCGAATGGAAGAAACCAAGGTCCAAAGCGAGAGAAATCCAGAGCAAGAAAGATC 5280
    |||
Db 5340 GTCGAATGGAAGAAACCAAGGTCCAAAGCGAGAGAAATCCAGAGCAAGAAAGATC 5399
QY 5281 TTGAGGGGGCTAGAAATCTGTTCTATGGCCCTTACACACATGCCACAGATCAACTG 5340
    |||
Db 5400 TTGAGGGGGCTAGAAATCTGTTCTATGGCCCTTACACACATGCCACAGATCAACTG 5459
QY 5341 GAATGATGTTAGAGTGTGTGCTGCTGTTGTTGAGAGAGTTTATCATCTACACCTT 5400
    |||
Db 5460 GAATGATGTTAGAGTGTGTGCTGCTGTTGTTGAGAGAGTTTATCATCTACACCTT 5519
QY 5401 GGCACAGGTGTCCACCAATTGTGTGTGCAACCCAGATGCTTGAGACAGAGCAATG 5460

```

```

Db 5520 GGCACAGGTGTCCACCAATTGTGTGCTGCTGTTGAGAGAGTTGCTGAGAGCAATG 5579
    |||
QY 5461 TTCCATGCATTTGGGCGAGATGTGTGAGGACACTTGTGTGACCCAGAGTGGTGTGAC 5520
    |||
Db 5580 TTCCATGCATTTGGGCGAGATGTGTGAGGACACTTGTGTGACCCAGAGTGGTGTGAC 5639
    |||
QY 5521 AGTGTACACTCTACAGTCCAGAGAGCTGAGACACTACCTGATATACCCAGATCC 5580
    |||
Db 5640 AGTGTACACTCTACAGTCCAGAGAGCTGAGACACTACCTGATATACCCAGATCC 5699
    |||
QY 5581 AGCCACTAC 5589
    |||
Db 5700 AGCCACTAC 5708

RESULT 14
AAT32601
ID AAT32601 standard; cDNA; 5914 BP.
XX
AC AAT32601;
XX
DT 19-NOV-1996 (first entry)
XX
DE BRCA1, breast and ovarian cancer susceptibility gene.
XX
KW BRCA1; breast cancer; ovary cancer; predisposing gene; diagnosis;
XX susceptibility gene; prognosis; gene therapy; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 120..5711
FT /*tag= a
FT /product= BRCA1 protein
XX
XX MO9605308-A1.
XX
PD 22-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10220.
XX
XX 07-JUN-1995; 95US-0488011.
XX 12-AUG-1994; 94US-0289221.
XX 02-SEP-1994; 94US-0300266.
XX 16-SEP-1994; 94US-0308104.
XX 29-NOV-1994; 94US-0348824.
XX 24-MAR-1995; 95US-0409305.
XX 07-JUN-1995; 95US-0483554.
XX 07-JUN-1995; 95US-0487002.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (UTAH ) UNIV UTAH RES FOUND.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Futreal PA, Goldgar DE, Harshtman KD, Kamb A, Miki Y;
PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;
PI Wiseman RW;
XX
DR WPI; 1996-139704/14.
DR P-PDB; AAR97128.
XX
XX
PT New method for diagnosing a predisposition to breast and ovarian
PT cancer - by detecting a germline alteration in the BRCA1 gene or
PT gene regulatory sequence; for gene therapy and to screen for drugs
XX
PS Claim 4; Page 108-117; 200pp; English.
XX
CC This is the nucleotide sequence of the breast and ovarian cancer
CC susceptibility gene, BRCA1. Four kindred families provided genetic
CC evidence for localisation of BRCA1 to a sufficiently small region for
CC the application of positional cloning strategies. A detailed map of
CC transcripts was developed for the region of 17q21 between D17S1321 and

```





```

Db 4260 GACCTGTCAGGGCTATCTCTCTCAGAGTGACATTTTAAACCACTGACGAGAGGATACCATG 4319
QY 4201 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAGCTGTGTAGAACAG 4260
Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAGCTGTGTAGAACAG 4379
QY 4261 CATGGAGCCAGCCTTCTTAACAGCTACCTTCCATCATAGTAGTACTTCTGCTTGGAG 4320
Db 4380 CATGGAGCCAGCCTTCTTAACAGCTACCTTCCATCATAGTAGTACTTCTGCTTGGAG 4439
QY 4321 GACCTGCGAAATCCAGAAACACATCATAGAAAGCAATATTAATCTTACAGAAAGT 4380
Db 4440 GACCTGCGAAATCCAGAAACACATCATAGAAAGCAATATTAATCTTACAGAAAGT 4499
QY 4381 AGTGAATACCTATTAAGCCAGATCCAGAAAGCCTTCTGTCAGCAAGTTTGAGGTCT 4440
Db 4500 AGTGAATACCTATTAAGCCAGATCCAGAAAGCCTTCTGTCAGCAAGTTTGAGGTCT 4559
QY 4441 GCAGATAGTTCTACAGTAAATAAAGAAACAGAGGTGAAAGTCAATCCCTTTTAA 4500
Db 4560 GCAGATAGTTCTACAGTAAATAAAGAAACAGAGGTGAAAGTCAATCCCTTTTAA 4619
QY 4501 TGGCCATCATTTAGATAGTGTGTACATGACAGATGCTCTGAGAGTCTTCAAGATAGA 4560
Db 4620 TGGCCATCATTTAGATAGTGTGTACATGACAGATGCTCTGAGAGTCTTCAAGATAGA 4679
QY 4561 AACTACCATCTCAAGAGAGCTCATTAGTGTGTAGTGTGAGAGCAACAGCTGGAA 4620
Db 4680 AACTACCATCTCAAGAGAGCTCATTAGTGTGTAGTGTGAGAGCAACAGCTGGAA 4739
QY 4621 GAGTCTGGGCCACAGCATTTGACGGAACATCTTACTTGGCAAGCAATCTAGAGGA 4680
Db 4740 GAGTCTGGGCCACAGCATTTGACGGAACATCTTACTTGGCAAGCAATCTAGAGGA 4799
QY 4681 ACCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGAGACCTGATGATCTCTCT 4740
Db 4800 ACCCTTACCTGGAATCTGGAATCAGCCTCTCTGATGAGACCTGATGATCTCTCTCT 4859
QY 4741 GAAGACAGAGCCCGACAGATCAGTCTGTGGCAGACATACATCTTCAACCTCTGATG 4800
Db 4860 GAAGACAGAGCCCGACAGATCAGTCTGTGGCAGACATACATCTTCAACCTCTGATG 4919
QY 4801 AAAGTTCCCAATTTGAAGTTGCAATCTGCCAGAGTCCAGCTCTGCTCACTACT 4860
Db 4920 AAAGTTCCCAATTTGAAGTTGCAATCTGCCAGAGTCCAGCTCTGCTCACTACT 4979
QY 4861 GATACCTGGGTATTAATGCAATGGAAGAGTGTGAGAGGAGAGCAAGATGACA 4920
Db 4980 GATACCTGGGTATTAATGCAATGGAAGAGTGTGAGAGGAGAGCAAGATGACA 5039
QY 4921 GCTTCACAGAAAGGTCACAAAGAAATGTCATGGTGTCTGGCTGACCCAGAA 4980
Db 5040 GCTTCACAGAAAGGTCACAAAGAAATGTCATGGTGTCTGGCTGACCCAGAA 5099
QY 4981 GAATTTATGCTGTGTACAGTTTGCAGAAACACACATCCTTTAACTAATCTAAT 5040
Db 5100 GAATTTATGCTGTGTACAGTTTGCAGAAACACACATCCTTTAACTAATCTAAT 5159
QY 5041 ACTGAAGAGTACTGATGTTGTATGAAGAACAGATGAGTGTGTGTGAACGACA 5100
Db 5160 ACTGAAGAGTACTGATGTTGTATGAAGAACAGATGAGTGTGTGTGAACGACA 5219
QY 5101 CTGAATATTTTCTAGGAATTCGCGAGAGAAATGGTGTAGTATTTCTGGTGACC 5160
Db 5220 CTGAATATTTTCTAGGAATTCGCGAGAGAAATGGTGTAGTATTTCTGGTGACC 5279
QY 5161 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGAGATGTC 5220
Db 5280 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGAGATGTC 5339
QY 5221 GTCAATGGAAGAACCAACCAAGTCCAAAGCGAGCAAGAGAAATCCCAAGCAAGAAAGT 5280
Db 5340 GTCAATGGAAGAACCAACCAAGTCCAAAGCGAGCAAGAGAAATCCCAAGCAAGAAAGT 5399

```

```

QY 5281 TTCAGGGGGCTAGAATCTGTTGCTATGAGCCCTTACCAACATGCCAGATCAACTG 5340
Db 5400 TTCAGGGGGCTAGAATCTGTTGCTATGAGCCCTTACCAACATGCCAGATCAACTG 5459
QY 5341 GAATGATGTGTACAGCTGTGTGTCTTGTGTGTGAAGAGCTTTTCAATTCACCTT 5400
Db 5460 GAATGATGTGTACAGCTGTGTGTCTTGTGTGTGAAGAGCTTTTCAATTCACCTT 5519
QY 5401 GGCACAGGTGTCCACCAATGTGTGTGTGAAGCAGATGCTTGACAGAGAGCAATGGC 5460
Db 5520 GGCACAGGTGTCCACCAATGTGTGTGTGTGAAGCAGATGCTTGACAGAGAGCAATGGC 5579
QY 5461 TTCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGTACCCGAGAGTGGTGTGGAC 5520
Db 5580 TTCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGTACCCGAGAGTGGTGTGGAC 5639
QY 5521 AGTGTAGCACTTACCAAGTGTGTGAGCAGAGCTGTGACACCTTACCTGATATCCCAATCCCCAC 5580
Db 5640 AGTGTAGCACTTACCAAGTGTGTGAGCAGAGCTGTGACACCTTACCTGATATCCCAATCCCCAC 5699
QY 5581 AGCCACTAC 5589
Db 5700 AGCCACTAC 5708

RESULT 15
ID AAT17438 standard; cDNA: 5914 BP.
AC AAT17438:
XX
AC
XX
DT 30-SEP-1996 (first entry)
XX
DE BRCA1 coding sequence.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds...
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 120..5711
XX FT /*tag= a
XX FT /product= BRCA1
XX
XX W09605306-A2.
XX
XX PD 22-FEB-1996.
XX
XX PE 11-AUG-1995: 95MO-US10202.
XX
XX PR 07-JUN-1995: 95US-0483553.
XX PR 12-AUG-1994: 94US-0289221.
XX PR 02-SEP-1994: 94US-0300266.
XX PR 16-SEP-1994: 94US-0308104.
XX PR 29-NOV-1994: 94US-0348824.
XX PR 24-MAR-1995: 95US-0409305.
XX PR 07-JUN-1995: 95US-0480784.
XX
XX PA (MYRI-) MYRIAD GENETICS INC.
XX PA (CANC-) CANCER INST.
XX PA (RECH-) CENT RECH DU CHUL.
XX
XX PI Durocher F, Eml M, Nakamura Y, Simard J, Shattuck-Eidens DM:
XX
XX WPI: 1996-139702/14.
XX DR P-PSDB; AAR81481.
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
XX gene - for diagnosis and therapy of human breast and ovarian cancer
XX and for diagnosing pre-disposition to these cancers

```

XX Claim 1; Page 119-128; 218bp; English.

PS This sequence represents the cDNA of the human breast and ovarian cancer  
 CC predisposing gene (BRCA1). Proteins encoded by mutations of this  
 CC sequence (see AAT17439-117453 and AAT17455-117529) can be used as  
 CC immunogens for antibody production. An additional BRCA1 mutant, with a T  
 CC to G mutation and 59 nucleotide insert (encoding a protein which  
 CC terminates at residue 75) is also claimed, however the exact location of  
 CC the mutations is not specified so the sequence could not be created. The  
 CC mutant BRCA1 genes have at least 1 mutation or polymorphism in comparison  
 CC to this sequence. By detecting a germline alteration in this gene, a  
 CC predisposition for breast and ovarian cancer can be diagnosed. In one  
 CC method, BRCA1 mRNA isolated from a tissue sample from a subject has a  
 CC probe, corresponding to a fragment of this sequence (or an  
 CC allele-specific probe for a mutation of this sequence), added to it.  
 CC The conditions allow for hybridisation of the probe to the mRNA, and any  
 CC hybridisation which occurs is detected. Alternatively the BRCA1 gene in  
 CC the tissue sample is isolated, and a shift in electrophoretic mobility  
 CC of single stranded DNA from the sample on a non-denaturing  
 CC polyacrylamide gel indicates a mutation. These methods of detection can  
 CC also diagnose a lesion neoplasia associated with the BRCA1 locus. The  
 CC methods may be used in gene therapy, protein replacement therapy and  
 CC protein mimetics, and may be used to screen for drugs in cancer therapy.

XX Sequence 5914 BP; 2006 A; 1156 C; 1316 G; 1436 T; 0 other;

Query Match 99.9%; Score 5585.8; DB 17; Length 5914;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGATTTATCTGCTCTGCTGCTGGAAGATGACAAATGCTATTATGCTATGAGAAA 60  
 Db ATGGATTTATCTGCTCTGCTGCTGGAAGATGACAAATGCTATTATGCTATGAGAAA 179  
 QY 61 ATCTTGAAGTGTCCATCTCTGCTGCTGATGATGACAACTGTCTCCAAAGTGTGAC 120  
 Db ATCTTGAAGTGTCCATCTCTGCTGCTGATGATGACAACTGTCTCCAAAGTGTGAC 239  
 QY 121 CACATATTTTGCATTTTGCATGCTGAAACCTCTGACGAGAAAGGCTTTCACAG 180  
 Db CACATATTTTGCATTTTGCATGCTGAAACCTCTGACGAGAAAGGCTTTCACAG 299  
 QY 240 CACATATTTTGCATTTTGCATGCTGAAACCTCTGACGAGAAAGGCTTTCACAG 299  
 Db CACATATTTTGCATTTTGCATGCTGAAACCTCTGACGAGAAAGGCTTTCACAG 359  
 QY 181 TGTCTTTATGATGATGATATATACCAAAAGGCTTACAAAGATGAGATTTAGT 240  
 Db TGTCTTTATGATGATGATATATACCAAAAGGCTTACAAAGATGAGATTTAGT 359  
 QY 300 TGTCTTTATGATGATGATATATACCAAAAGGCTTACAAAGATGAGATTTAGT 359  
 Db TGTCTTTATGATGATGATATATACCAAAAGGCTTACAAAGATGAGATTTAGT 419  
 QY 241 CAACTTGTGAGAGCTATTTGAAATCATTTTGTCTTTCAGCTTGACACAGTTGAG 300  
 Db CAACTTGTGAGAGCTATTTGAAATCATTTTGTCTTTCAGCTTGACACAGTTGAG 419  
 QY 360 CAACTTGTGAGAGCTATTTGAAATCATTTTGTCTTTCAGCTTGACACAGTTGAG 419  
 Db CAACTTGTGAGAGCTATTTGAAATCATTTTGTCTTTCAGCTTGACACAGTTGAG 360  
 QY 301 TATGCAACAGCATATTTTGCAGAAAGGAAATTAACCTTCTTCACTTAAAGAT 360  
 Db TATGCAACAGCATATTTTGCAGAAAGGAAATTAACCTTCTTCACTTAAAGAT 479  
 QY 420 TATGCAACAGCATATTTTGCAGAAAGGAAATTAACCTTCTTCACTTAAAGAT 479  
 Db TATGCAACAGCATATTTTGCAGAAAGGAAATTAACCTTCTTCACTTAAAGAT 539  
 QY 480 GAAGTTTATATATCAAAAGATAGGCTACAGAAACCGTGCCAAAGACTTCTACAGAGT 539  
 Db GAAGTTTATATATCAAAAGATAGGCTACAGAAACCGTGCCAAAGACTTCTACAGAGT 480  
 QY 421 GAACCCGAAATATCTTCTTGCAGAAACAGCTCTAGTGTCCAACTCTCTAACTTGG 480  
 Db GAACCCGAAATATCTTCTTGCAGAAACAGCTCTAGTGTCCAACTCTCTAACTTGG 599  
 QY 540 GAACCCGAAATATCTTCTTGCAGAAACAGCTCTAGTGTCCAACTCTCTAACTTGG 599  
 Db GAACCCGAAATATCTTCTTGCAGAAACAGCTCTAGTGTCCAACTCTCTAACTTGG 540  
 QY 481 ACTGTGAGACCTGAGGAGCAAGGAGGATCAACCTCAAAAGAGCTGTCTCAAT 540  
 Db ACTGTGAGACCTGAGGAGCAAGGAGGATCAACCTCAAAAGAGCTGTCTCAAT 659  
 QY 600 ACTGTGAGACCTGAGGAGCAAGGAGGATCAACCTCAAAAGAGCTGTCTCAAT 659  
 Db ACTGTGAGACCTGAGGAGCAAGGAGGATCAACCTCAAAAGAGCTGTCTCAAT 600  
 QY 541 GAATGGAGTCTGATCTTCTGAGATACCGTTAATAGGCAACTATTGCACTGTGG 600  
 Db GAATGGAGTCTGATCTTCTGAGATACCGTTAATAGGCAACTATTGCACTGTGG 719  
 QY 660 GAATGGAGTCTGATCTTCTGAGATACCGTTAATAGGCAACTATTGCACTGTGG 719  
 Db GAATGGAGTCTGATCTTCTGAGATACCGTTAATAGGCAACTATTGCACTGTGG 601  
 QY 601 GATCAAGATTTGTACAAATCACCCCTCAAGAGACGAGGATGAATCACTTGTCTT 660  
 Db GATCAAGATTTGTACAAATCACCCCTCAAGAGACGAGGATGAATCACTTGTCTT 1859

Db 720 GATCAAGATTTGTACAAATCACCCCTCAAGAGACGAGGATGAATCACTTGTCTT 779  
 QY 661 GCAAAAAGGCTCTTGTGATTTTCTGAGACGATGTAAACAATACTGAATCATCA 720  
 Db GCAAAAAGGCTCTTGTGATTTTCTGAGACGATGTAAACAATACTGAATCATCA 839  
 QY 721 CCCAGTAATATGATTTGAAACACCACTGAGAACGCTGACAGGATCCAGAAAG 780  
 Db CCCAGTAATATGATTTGAAACACCACTGAGAACGCTGACAGGATCCAGAAAG 899  
 QY 840 CCCAGTAATATGATTTGAAACACCACTGAGAACGCTGACAGGATCCAGAAAG 899  
 Db CCCAGTAATATGATTTGAAACACCACTGAGAACGCTGACAGGATCCAGAAAG 959  
 QY 781 TATCAGGATGATTTCTTTCAAACTTGCATGTGAGACCATGTGGCAAAATCTATGCC 840  
 Db TATCAGGATGATTTCTTTCAAACTTGCATGTGAGACCATGTGGCAAAATCTATGCC 959  
 QY 900 TATCAGGATGATTTCTTTCAAACTTGCATGTGAGACCATGTGGCAAAATCTATGCC 959  
 Db TATCAGGATGATTTCTTTCAAACTTGCATGTGAGACCATGTGGCAAAATCTATGCC 841  
 QY 841 AGCTCATTTACAGATGAGAACACCACTTATTTCTACATAAAGACGAATGATGAGA 900  
 Db AGCTCATTTACAGATGAGAACACCACTTATTTCTACATAAAGACGAATGATGAGA 1019  
 QY 960 AGCTCATTTACAGATGAGAACACCACTTATTTCTACATAAAGACGAATGATGAGA 1019  
 Db AGCTCATTTACAGATGAGAACACCACTTATTTCTACATAAAGACGAATGATGAGA 901  
 QY 901 AAGGCTGAATTTCTGTATTAAGCAAAACAGCTGCTTACAGAGGACCAATACAGA 960  
 Db AAGGCTGAATTTCTGTATTAAGCAAAACAGCTGCTTACAGAGGACCAATACAGA 1079  
 QY 1020 AAGGCTGAATTTCTGTATTAAGCAAAACAGCTGCTTACAGAGGACCAATACAGA 1079  
 Db AAGGCTGAATTTCTGTATTAAGCAAAACAGCTGCTTACAGAGGACCAATACAGA 1080  
 QY 961 TGGGCTGGAAGTAAAGAAACATGATATGATAGCGGACCTCCAGCAGAAAAAGGTA 1139  
 Db TGGGCTGGAAGTAAAGAAACATGATATGATAGCGGACCTCCAGCAGAAAAAGGTA 1021  
 QY 1021 GATCTGAATGCTGATCCCTGCTGAGAGAAAGAAATGATTAAGCAAACTGCATGC 1080  
 Db GATCTGAATGCTGATCCCTGCTGAGAGAAAGAAATGATTAAGCAAACTGCATGC 1140  
 QY 1140 GATCTGAATGCTGATCCCTGCTGAGAGAAAGAAATGATTAAGCAAACTGCATGC 1199  
 Db GATCTGAATGCTGATCCCTGCTGAGAGAAAGAAATGATTAAGCAAACTGCATGC 1081  
 QY 1081 TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTTAATAGCAGATTTCAG 1140  
 Db TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTTAATAGCAGATTTCAG 1200  
 QY 1200 TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTTAATAGCAGATTTCAG 1259  
 Db TCAGAGAAATCTAGAGATCTAGAGATGTTCTTGTGATTAACACTTAATAGCAGATTTCAG 1141  
 QY 1141 AAAGTTAATGAGGCTTTTCCAGAAATGATGAACCTGTAGGTTCTATGATCATGAT 1200  
 Db AAAGTTAATGAGGCTTTTCCAGAAATGATGAACCTGTAGGTTCTATGATCATGAT 1260  
 QY 1260 AAAGTTAATGAGGCTTTTCCAGAAATGATGAACCTGTAGGTTCTATGATCATGAT 1319  
 Db AAAGTTAATGAGGCTTTTCCAGAAATGATGAACCTGTAGGTTCTATGATCATGAT 1201  
 QY 1201 GGGGAGCTGGAATCAATGCAAGAGTGAAGTGAATGTTGAGCTGTTTAATGAGTATGAT 1260  
 Db GGGGAGCTGGAATCAATGCAAGAGTGAAGTGAATGTTGAGCTGTTTAATGAGTATGAT 1320  
 QY 1320 GGGGAGCTGGAATCAATGCAAGAGTGAAGTGAATGTTGAGCTGTTTAATGAGTATGAT 1379  
 Db GGGGAGCTGGAATCAATGCAAGAGTGAAGTGAATGTTGAGCTGTTTAATGAGTATGAT 1261  
 QY 1261 GAATATTTCTGTTCTTCAAGAAATATGACTTACTGCGCAGTATCTCATGAGCTTTA 1320  
 Db GAATATTTCTGTTCTTCAAGAAATATGACTTACTGCGCAGTATCTCATGAGCTTTA 1380  
 QY 1380 GAATATTTCTGTTCTTCAAGAAATATGACTTACTGCGCAGTATCTCATGAGCTTTA 1439  
 Db GAATATTTCTGTTCTTCAAGAAATATGACTTACTGCGCAGTATCTCATGAGCTTTA 1321  
 QY 1321 ATATGTAAGAGTGAAGAGTTCCTCAATCAATCAGTGAAGATATATTTGAGCAAAATA 1380  
 Db ATATGTAAGAGTGAAGAGTTCCTCAATCAATCAGTGAAGATATATTTGAGCAAAATA 1440  
 QY 1440 ATATGTAAGAGTGAAGAGTTCCTCAATCAATCAGTGAAGATATATTTGAGCAAAATA 1499  
 Db ATATGTAAGAGTGAAGAGTTCCTCAATCAATCAGTGAAGATATATTTGAGCAAAATA 1381  
 QY 1381 TTTGGGAAAACTATGGAAGGAGGAGGCTCCCACTTAAAGCATGTAACTGAAAT 1440  
 Db TTTGGGAAAACTATGGAAGGAGGAGGCTCCCACTTAAAGCATGTAACTGAAAT 1500  
 QY 1500 TTTGGGAAAACTATGGAAGGAGGAGGCTCCCACTTAAAGCATGTAACTGAAAT 1559  
 Db TTTGGGAAAACTATGGAAGGAGGAGGCTCCCACTTAAAGCATGTAACTGAAAT 1441  
 QY 1441 CTAAATTAAGAGACATTTTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAAT 1500  
 Db CTAAATTAAGAGACATTTTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAAT 1560  
 QY 1560 CTAAATTAAGAGACATTTTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAAT 1619  
 Db CTAAATTAAGAGACATTTTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAAT 1501  
 QY 1501 AAATTAAGAGCTTAAGAGAGACCTACATCAGGCTTCTATCTGAGATTTTATCAAGAAA 1560  
 Db AAATTAAGAGCTTAAGAGAGACCTACATCAGGCTTCTATCTGAGATTTTATCAAGAAA 1620  
 QY 1620 AAATTAAGAGCTTAAGAGAGACCTACATCAGGCTTCTATCTGAGATTTTATCAAGAAA 1679  
 Db AAATTAAGAGCTTAAGAGAGACCTACATCAGGCTTCTATCTGAGATTTTATCAAGAAA 1561  
 QY 1561 GCAGATTTGGCAGTTGAAAAGACCTCTGAAATGATTAATCAAGGAGCTTAACCAAGGAG 1620  
 Db GCAGATTTGGCAGTTGAAAAGACCTCTGAAATGATTAATCAAGGAGCTTAACCAAGGAG 1680  
 QY 1680 GCAGATTTGGCAGTTGAAAAGACCTCTGAAATGATTAATCAAGGAGCTTAACCAAGGAG 1739  
 Db GCAGATTTGGCAGTTGAAAAGACCTCTGAAATGATTAATCAAGGAGCTTAACCAAGGAG 1621  
 QY 1621 CAGAAATGCTCAAGTGAATTAATTAATAGTGTCTATGAGATTAACCAAGGCTGAT 1680  
 Db CAGAAATGCTCAAGTGAATTAATTAATAGTGTCTATGAGATTAACCAAGGCTGAT 1740  
 QY 1740 CAGAAATGCTCAAGTGAATTAATTAATAGTGTCTATGAGATTAACCAAGGCTGAT 1799  
 Db CAGAAATGCTCAAGTGAATTAATTAATAGTGTCTATGAGATTAACCAAGGCTGAT 1681  
 QY 1681 TCTATTCAAGATGAGAAAAATCTTAACCAATTAATGATCTGAGAAAAAGATCTGCTTTC 1740  
 Db TCTATTCAAGATGAGAAAAATCTTAACCAATTAATGATCTGAGAAAAAGATCTGCTTTC 1800  
 QY 1800 TCTATTCAAGATGAGAAAAATCTTAACCAATTAATGATCTGAGAAAAAGATCTGCTTTC 1859  
 Db TCTATTCAAGATGAGAAAAATCTTAACCAATTAATGATCTGAGAAAAAGATCTGCTTTC



1741 AAAACGAAAGCTGAACCTATTAAGCAGATATAGCAATATGGAACCTGCAATTAAATATC 1800  
1860 AAAAGCAAGAGCTAAACCTATTAAGCAGATATAGCAATATGGAACCTGCAATTAAATATC 1919  
1801 CACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAAAGTCTTACAGGCAATAT 1860  
1920 CACAATTCAAAAGCAGCTTAAAGAAATAGGCTGAGAGAGAAAGTCTTACAGGCAATAT 1979  
1861 CATTGCTTGAACCTAGTACATAGCAATCTAAGCCACCTATGCTAGTGAATTCGA 1920  
1980 CATGCTTGAACCTAGTACATAGCAATCTAAGCCACCTATGCTAGTGAATTCGA 2039  
1921 ATTGATAGTGTCTTCTAGCAGTGAAGATTAAGAAAAAAAGTACAACCAATGCCAGTC 1980  
2040 ATTGATAGTGTCTTCTAGCAGTGAAGATTAAGAAAAAAAGTACAACCAATGCCAGTC 2099  
1981 AGGCACAGCAGAAACCTACACATCTATGAAAGGTAAGAAACCTGCAACTGAGCCAGAA 2040  
2100 AGGCACAGCAGAAACCTACACATCTATGAAAGGTAAGAAACCTGCAACTGAGCCAGAA 2159  
2041 AGTAAACAGCAAAATGAACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTG 2100  
2160 AGTAAACAGCAAAATGAACAGACAGTAAAGACATGACAGTATCTTCCAGAGCTG 2219  
2101 AAGTTAAACAATGCACTGCTTCTTCTTACTAAGTGTTCAAATACAGTGAATTTAAAGAA 2160  
2220 AAGTTAAACAATGCACTGCTGCTTCTTCTTACTAAGTGTTCAAATACAGTGAATTTAAAGAA 2279  
2161 TTTGTCATCTTACCTTCCAGAGAGAGAGAGAGAGAGAACTAGAACAGTTAAAGTG 2220  
2280 TTTGTCATCTTACCTTCCAGAGAGAGAGAGAGAGAGAACTAGAACAGTTAAAGTG 2339  
2221 TCTAATTAATGCTGAAGACCCCAAGAGATCTATGTTAAGTGAAGAGGTTTGGCAACT 2280  
2340 TCTAATTAATGCTGAAGACCCCAAGAGATCTATGTTAAGTGAAGAGGTTTGGCAACT 2399  
2281 GAAAGATCTGAGAGAGTGAAGATTTATGTTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2340  
2400 GAAAGATCTGAGAGAGTGAAGATTTATGTTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2459  
2341 GAAAGATCTGAGAGAGTGAAGATTTATGTTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2400  
2460 GAAAGATCTGAGAGAGTGAAGATTTATGTTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2519  
2401 TGTGAGTCACTGTCAGACATTTGAAACCCCAAGAGACTAATTCATGTTTTCGCAAA 2460  
2520 TGTGAGTCACTGTCAGACATTTGAAACCCCAAGAGACTAATTCATGTTTTCGCAAA 2579  
2461 GATAATTAAGATGACACAGAGGCTTTAAGTATCCATTGGAGCATGAAGTTAACACAGT 2520  
2580 GATAATTAAGATGACACAGAGGCTTTAAGTATCCATTGGAGCATGAAGTTAACACAGT 2639  
2521 CGGGAAACAGAGATGAATGAAGAACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580  
2640 CGGGAAACAGAGATGAATGAAGAACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2699  
2581 TTCAAGTGTTCAAAGCGCAGTCAATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2640  
2700 TTCAAGTGTTCAAAGCGCAGTCAATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2759  
2641 GAATGTCAGACATTTCTGCGCAGTCTGCGCTTAAAGAAACAAAGTCCAAAGCTCACT 2700  
2760 GAATGTCAGACATTTCTGCGCAGTCTGCGCTTAAAGAAACAAAGTCCAAAGCTCACT 2819  
2701 TTTGATATGGAACAAAGAGAGAAATCAAGAGAAAGATGAGCTTAATATCAAGCTGTA 2760  
2820 TTTGATATGGAACAAAGAGAGAAATCAAGAGAAAGATGAGCTTAATATCAAGCTGTA 2879  
2761 CAGACAGTTAATATCAGTCAAGGCTTCTGTTGTTGTCAGAGAAATGAGCTTAATATCAAG 2820  
2880 CAGACAGTTAATATCAGTCAAGGCTTCTGTTGTTGTCAGAGAAATGAGCTTAATATCAAG 2939

2821 AATGCCAAATGATATCAAAAGAGGCTCTAGGTTTGTCTATCACTCAGTTCAGAGC 2880  
2940 AATGCCAAATGATATCAAAAGAGGCTCTAGGTTTGTCTATCACTCAGTTCAGAGC 2999  
2881 AACGAACCTGAGACTATTAATCCAAATTAACATGAGACTTTTCAAAACCCATATCTGTA 2940  
3000 AACGAACCTGAGACTATTAATCCAAATTAACATGAGACTTTTCAAAACCCATATCTGTA 3059  
2941 CCACCACTTTTCCCTCAGTCAATTTGTTAAACTAAATGTAACAAAATCTGCTAGAG 3000  
3060 CCACCACTTTTCCCTCAGTCAATTTGTTAAACTAAATGTAACAAAATCTGCTAGAG 3119  
3001 GAAACTTTGAGGAACTTCAATGTCACCTGAAAGAGAAATGGAAGAAAGTGAACATTTCA 3060  
3120 GAAACTTTGAGGAACTTCAATGTCACCTGAAAGAGAAATGGAAGAAAGTGAACATTTCA 3179  
3061 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAGAAATGTTTAAAGAGCCAGC 3120  
3180 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAGAAATGTTTAAAGAGCCAGC 3239  
3121 TCAGCAATTAATGAATGAGTGAAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3180  
3240 TCAGCAATTAATGAATGAGTGAAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3299  
3181 ATAGGTTCCAGTATCAAAACATTCAGAGAGAACTAGTGAAGAAACAGAGGCCAAATTTG 3240  
3300 ATAGGTTCCAGTATCAAAACATTCAGAGAGAACTAGTGAAGAAACAGAGGCCAAATTTG 3359  
3241 AATGCTATCTTATGATTAAGGCTTTTGCACCTGAGTCTATTAACAAAGTCTTCTGGA 3300  
3360 AATGCTATCTTATGATTAAGGCTTTTGCACCTGAGTCTATTAACAAAGTCTTCTGGA 3419  
3301 AGTAAATGTAACATCTGTAATTAAGAAAGAAAGTGAAGAAAGTGAAGTGAAGTGAAGTGAAG 3360  
3420 AGTAAATGTAACATCTGTAATTAAGAAAGAAAGTGAAGAAAGTGAAGTGAAGTGAAGTGAAG 3479  
3361 AATACAGATTTCTCCATATCTGATTTCAATTAAGTGAAGAAAGTGAAGTGAAGTGAAGTGAAG 3420  
3480 AATACAGATTTCTCCATATCTGATTTCAATTAAGTGAAGAAAGTGAAGTGAAGTGAAGTGAAG 3539  
3421 CATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGATGATGATGATGATGAT 3480  
3540 CATGATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGATGATGATGATGATGAT 3599  
3481 GAAAGATCTAGTGTTCGAAATGACATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3540  
3600 GAAAGATCTAGTGTTCGAAATGACATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3659  
3541 GTCCAGAAAGGAGACCTTAGCAGAGTCTAGCCCTTTCACCCATTAACATTTGGCTCAG 3600  
3660 GTCCAGAAAGGAGACCTTAGCAGAGTCTAGCCCTTTCACCCATTAACATTTGGCTCAG 3719  
3601 GGTTCACGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3660  
3720 GGTTCACGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3779  
3661 GAAGAGCTTCCCTGCTTCCACACATTTGATTTGGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3720  
3780 GAAGAGCTTCCCTGCTTCCACACATTTGATTTGGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3839  
3721 TCTACTAGGCAATGACAGGCTTGAACGAGTCTGCTTAAGAAACAGAGAGGAGAAATTTA 3780  
3840 TCTACTAGGCAATGACAGGCTTGAACGAGTCTGCTTAAGAAACAGAGAGGAGAAATTTA 3899  
3781 TTATCATTTGAGAAATAGCTTTAAATGACTGCAATTAACAGGTAATTTGGCAAGAGCATCT 3840  
3900 TTATCATTTGAGAAATAGCTTTAAATGACTGCAATTAACAGGTAATTTGGCAAGAGCATCT 3959  
3841 CAGGAACATCACTTATGAG 3900  
3960 CAGGAACATCACTTATGAG 4019  
3901 AGTGAATGGAAGACTGCTGCAATTAACAAACACAGAGATCCTTCTTGAATGGTCT 3960







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 16:19:58 ; Search time 7050.55 Seconds  
(without alignments)  
12838.232 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_120\_5708

Perfect score: 5589  
Sequence: 1 ATGATTTATCTGCTCTCTG.....AGATCCCAACAGACACTAC 5589

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:\*  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1960.8	35.1	2101	11	BC012577 Homo sapi
2	1955	35.0	2090	11	BC030969 Homo sapi
3	861.4	15.4	962	14	B0068830 AGENCOURT
4	843	15.1	1089	13	BM452288 AGENCOURT
5	811.8	14.5	845	9	AU122476
6	761.8	13.6	783	9	AU142729

7	740.4	13.2	743	12	BG681276
8	688	12.3	702	12	AU125312
9	686.6	12.3	739	12	BF508987
10	686.4	12.3	897	14	BM800251
11	666.6	11.9	899	14	BQ422380
12	657	11.8	900	12	BF791668
13	619.2	11.1	921	12	BG178466
14	616.2	11.0	878	14	BQ215100
15	603.8	10.8	630	9	A1992040
16	592.6	10.6	800	12	BF794879
17	591.4	10.6	627	9	A1589028
18	591	10.6	724	9	A1915085
19	590	10.6	906	14	BQ878445
20	590	10.6	909	14	BQ676829
21	590	10.6	933	14	BQ683955
22	587	10.5	747	12	BG777447
23	583	10.4	947	14	BQ679749
24	578	10.3	910	14	BQ677666
25	575	10.3	602	10	BE043993
26	561.6	10.0	938	12	BG257190
27	560.4	10.0	675	9	AL704228
28	545.2	9.8	628	13	BM042282
29	536.8	9.6	638	10	BE264293
30	531.6	9.5	563	10	AW295197
31	530.2	9.5	585	10	AW968546
32	518	9.3	666	10	AW968720
33	499.8	8.9	509	10	AW504244
34	492.8	8.8	518	9	AA804632
35	491.8	8.8	523	9	AA702344
36	491.2	8.8	509	10	BE018878
37	479.4	8.6	1220	14	BQ214737
38	472.2	8.4	487	9	AA812019
39	450.6	8.1	575	10	BE564528
40	449.8	8.0	453	9	A1684595
41	444	7.9	444	9	AA486004
42	437.4	7.8	488	9	AA814998
43	421.8	7.5	429	9	AA484941
44	420.6	7.5	459	9	A1040685
45	418.2	7.5	443	14	BQ308670

## ALIGNMENTS

RESULT 1  
BC012577  
LOCUS BC012577 2101 bp mRNA Linear HTC 20-AUG-2001  
DEFINITION Homo sapiens, clone IMAGE:3996558, mRNA.  
ACCESSION BC012577  
VERSION BC012577.1 GI:15214876  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE  
1 (bases 1 to 2101)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-femail.nih.gov](mailto:cgaps-femail.nih.gov)  
Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNU)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRRL Plate: 19 Row: a Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6552298  
 This clone has the following problem: incomplete processing.

## FEATURES

Location/Qualifiers

1.. 2101

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:399658"

/tissue\_type="Bladder, carcinoma"

/clone\_lib="NIH\_MGC\_53"

/lab\_host="DH10B"

/note="Vector: pDNR-LIB"

BASE COUNT 756 a 385 c 452 g 508 t

## ORIGIN

Query Match 35.1%; Score 1960.8; DB 11; Length 2101;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1965; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 ATGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCAATATGCTATGCGAAG 60  
 |||||  
 124 ATGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCAATATGCTATGCGAAG 183  
 |||||  
 61 ATCTAGAGTGTCCATCTGCTGAGTGATCAAGAACCTGTCCCAAAAGTGAC 120  
 |||||  
 184 ATCTAGAGTGTCCATCTGCTGAGTGATCAAGAACCTGTCCCAAAAGTGAC 243  
 |||||  
 121 CACATATTTTGAATTTTGCATGCTGGAACCTTCTACAGAGAAGAGGCGCTTCACAG 180  
 |||||  
 244 CACATATTTTGAATTTTGCATGCTGGAACCTTCTACAGAGAAGAGGCGCTTCACAG 303  
 |||||  
 181 TGTCTTTATGTAAGATATATTAACCAAAAGAGCTACAGAAAGTACAGATTTAGT 240  
 |||||  
 304 TGTCTTTATGTAAGATATATTAACCAAAAGAGCTACAGAAAGTACAGATTTAGT 353  
 |||||  
 241 CAATCTGTTGAAGACTATTTGAATCATTTTGTGCTTTCAGCTTACAGAGTTGGAG 300  
 |||||  
 364 CAATCTGTTGAAGACTATTTGAATCATTTTGTGCTTTCAGCTTACAGAGTTGGAG 423  
 |||||  
 301 TATGCAAAACAGCTATTAATTTTGCACAAAAGAAATTAAGTCTCTGACATCTAAAGAT 360  
 |||||  
 424 TATGCAAAACAGCTATTAATTTTGCACAAAAGAAATTAAGTCTCTGACATCTAAAGAT 483  
 |||||  
 361 GAAGTTTCTATCATCTCAAGATATGGGCTACAGAAACCGTCCAAAAGACTTCTACAGAGT 420  
 |||||  
 484 GAAGTTTCTATCATCTCAAGATATGGGCTACAGAAACCGTCCAAAAGACTTCTACAGAGT 543  
 |||||  
 421 GAACCCGAAAATCTCTCTTTCGAGAGAAACAGCTCAGTCACTCTTAACCTTGGG 480  
 |||||  
 544 GAACCCGAAAATCTCTCTTTCGAGAGAAACAGCTCAGTCACTCTTAACCTTGGG 603  
 |||||  
 481 ACTGTGAGAACTCTGAGGACAAAGCAGCGATACAACTCAAAAGAGCTCTGTCTACAT 540  
 |||||  
 604 ACTGTGAGAACTCTGAGGACAAAGCAGCGATACAACTCAAAAGAGCTCTGTCTACAT 663  
 |||||  
 541 GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATTAAGCAACTTATTGCACTGGGA 600  
 |||||  
 664 GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATTAAGCAACTTATTGCACTGGGA 723  
 |||||  
 601 GATCAAGATTTGTTCAAAATCAACCCCTCAAGGAAACAGGATGAATCACTGTTGATTC 660  
 |||||  
 724 GATCAAGATTTGTTCAAAATCAACCCCTCAAGGAAACAGGATGAATCACTGTTGATTC 783  
 |||||  
 661 GCAAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAACAAATACTAGATCATCA 720  
 |||||  
 784 GCAAAAAAGGCTGCTTGTGAATTTTCTGAGAGGATGTAACAAATACTAGATCATCA 843  
 |||||  
 721 CCCAGTAAATGATTTGAACACCACTGGAAGCGTGACGCTGAGAGGATCCAGAAAG 780  
 |||||

844 CCCAGTAAATGATTTGAACACCACTGGAAGCGTGAGTGAAGGATTCAGAAAAG 903  
 |||||  
 781 TATCAGAGTATCTCTCTTCAAACTGATGAGAGCATGTGGACAAATATCATGCC 840  
 |||||  
 904 TATCAGAGTATCTCTCTTCAAACTGATGAGAGCATGTGGACAAATATCATGCC 963  
 |||||  
 841 AGCTCATTTACAGATGAGAACAGCAGTTTATTTACTACTAAGACAGATGATGTAGA 900  
 |||||  
 964 AGCTCATTTACAGATGAGAACAGCAGTTTATTTACTACTAAGACAGATGATGTAGA 1023  
 |||||  
 901 AAGGCTGAATTTCTGATTAATAAGCAACACCGCTGCTTACAGAGAGCCCAATACGA 960  
 |||||  
 1024 AAGGCTGAATTTCTGATTAATAAGCAACACCGCTGCTTACAGAGAGCCCAATACGA 1083  
 |||||  
 961 TGGGCTGGAAGTAAAGAAACATGTAATGATGAGGAGCATCCAGACAGAAAAGGTA 1020  
 |||||  
 1084 TGGGCTGGAAGTAAAGAAACATGTAATGATGAGGAGCATCCAGACAGAAAAGGTA 1143  
 |||||  
 1021 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAGATGGAATTAACAGAACTGCCATGC 1080  
 |||||  
 1144 GATCTGAATGCTGATCCCTGCTGTGAGAGAAAAGATGGAATTAACAGAACTGCCATGC 1203  
 |||||  
 1081 TCAGAGATCTAGAGATACTGAGATGTTCTTGGATTAACCTAATTAAGCAGATTGAG 1140  
 |||||  
 1204 TCAGAGATCTAGAGATACTGAGATGTTCTTGGATTAACCTAATTAAGCAGATTGAG 1263  
 |||||  
 1141 AAGGTAATGAGAGGTTTCCAGAGTGATGATGATGATGATGATGATGATGATGATGAT 1200  
 |||||  
 1264 AAGGTAATGAGAGGTTTCCAGAGTGATGATGATGATGATGATGATGATGATGATGAT 1323  
 |||||  
 1201 GGGGAGTCTGAATCAATCCCAAGTAGCTGATGATGAGAGTTCTAATGAGGTAGAT 1260  
 |||||  
 1324 GGGGAGTCTGAATCAATCCCAAGTAGCTGATGATGATGATGATGATGATGATGATGAT 1383  
 |||||  
 1261 GAATATTCGTGTTCTTCAAGAGAAATAGACTTACGCGCAGAGTACCTATGAGGCTTTA 1320  
 |||||  
 1384 GAATATTCGTGTTCTTCAAGAGAAATAGACTTACGCGCAGAGTACCTATGAGGCTTTA 1443  
 |||||  
 1321 ATATGTAAGTGAAGAGTCTACTCCAAATCAGTAGAGTAATTAAGAGACAAATA 1380  
 |||||  
 1444 ATATGTAAGTGAAGAGTCTACTCCAAATCAGTAGAGTAATTAAGAGACAAATA 1503  
 |||||  
 1381 TTTGGGAAAACCTATCGAAGAAAGCAGCCTCCCACTTAAGCATGTAAGTAAAT 1440  
 |||||  
 1504 TTTGGGAAAACCTATCGAAGAAAGCAGCCTCCCACTTAAGCATGTAAGTAAAT 1563  
 |||||  
 1441 CTATATTATGAGGACTTTGTTACTGAGCCACAGATTAATCAAGAGCGTCCCTCAAT 1500  
 |||||  
 1564 CTATATTATGAGGACTTTGTTACTGAGCCACAGATTAATCAAGAGCGTCCCTCAAT 1623  
 |||||  
 1501 AATTTAAACCGTAAAGAGACCTACATCAGCGCTTCATCTGAGAGTTTATCAAGAAA 1560  
 |||||  
 1624 AATTTAAACCGTAAAGAGACCTACATCAGCGCTTCATCTGAGAGTTTATCAAGAAA 1683  
 |||||  
 1561 GCAGATTTGGCACTTCAAAAAGCTCTGAATGATTAATCAAGGAACTTAACCAAGGAG 1620  
 |||||  
 1684 GCAGATTTGGCACTTCAAAAAGCTCTGAATGATTAATCAAGGAACTTAACCAAGGAG 1743  
 |||||  
 1621 CAGATTTGGCACTTCAAAAAGCTCTGAATGATTAATCAAGGAACTTAACCAAGGAG 1680  
 |||||  
 1744 CAGATTTGGCACTTCAAAAAGCTCTGAATGATTAATCAAGGAACTTAACCAAGGAG 1803  
 |||||  
 1681 TCTATTCAGATGAGAAAATCTCTAACCATAATGATCACTGAGAAAAGATCTGCTTTC 1740  
 |||||  
 1804 TCTATTCAGATGAGAAAATCTCTAACCATAATGATCACTGAGAAAAGATCTGCTTTC 1863  
 |||||  
 1741 AAAAGCAAGCTGAACCTATAAGCAGAGTATTAAGCAATATGGAATCTGAATTAATATC 1800  
 |||||  
 1864 AAAAGCAAGCTGAACCTATAAGCAGAGTATTAAGCAATATGGAATCTGAATTAATATC 1923  
 |||||  
 1801 CACATTTCAAAAGCAGCTTAAGAAATAGGCTGAGAGGAAAGCTTTCTACAGGCAATAT 1860  
 |||||  
 1924 CACATTTCAAAAGCAGCTTAAGAAATAGGCTGAGAGGAAAGCTTTCTACAGGCAATAT 1983  
 |||||

QY 1861 CATGCCCTTGAAGTACTAGTACAGTAAGTAACTAAGCCACCTAATGTACTGTAATGCAA 1920  
|||||  
Db 1984 CATGCCCTTGAAGTACTAGTACAGTAAGTAACTAAGCCACCTAATGTACTGTAATGCAA 2043  
QY 1921 ATTGATAGTGTGTCTGTACAGTGAAGATPAAGAAAAAATAGCAACCAA 1972  
|||||  
Db 2044 ATTGATAGTGTGTCTGTACAGTGAAGATPAAGAAAAAATAGCAACCAA 2095  
|||||  
RESULT 2  
BC030969 2090 bp mRNA linear HTC 13-JUN-2002  
LOCUS Homo sapiens, similar to breast cancer 1, early onset, clone  
DEFINITION IMAGE:4804551, mRNA.  
ACCESSION BC030969  
VERSION BC030969.1 GI:21411299  
KEYWORDS HTC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2090)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: http://mgc.ncl.nih.gov  
REMARK Contact: MGC help desk  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTP/DRP  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: http://www.shgc.stanford.edu  
Contact: (Dickson, Mark) mcdexaxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
Series: IRAL Plate: 41 Row: P Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6552298  
This clone has the following problem: frame shifted.  
FEATURES  
Location/Qualifiers  
1..2090  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4804551"  
/tissue\_type="Prostate, adenocarcinoma."  
/lab\_host="NIH\_MGC\_60"  
/note="Vector: pDNR-LIB"  
BASE COUNT 754 a 385 c 444 g 507 t  
ORIGIN  
Query Match 35.0%; Score 1955; DB 11; Length 2090;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1961; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 ARGGATTTCCTGCTTGGCGTTGAAGAAGTAAATGTCTAATAGCTGACGAAA 60  
|||||  
Db 120 ATGGATTATCTGCTCTTGGCGTTGAAGAAGTAAATGTCTAATAGCTGACGAAA 179  
|||||  
QY 61 AACTTGAAGTCCCATCTGTCTGAGTGTATCAAGAACCTGTCCCAAGAGTGTGAC 120  
|||||  
Db 180 ATCTTGAAGTCCCATCTGTCTGAGTGTATCAAGAACCTGTCTCCCAAGAGTGTGAC 239  
|||||  
QY 121 CACATATTTTGCAGAAATTTTGCATGTGAACCTTCTCACCCAGAAAGAGGCGCTTCACAG 180  
|||||

|||||  
Db 240 CACATATTTTGCAGAAATTTTGCATGTGAACCTTCTCACCCAGAAAGAGGCGCTTCACAG 299  
|||||  
QY 181 TGTGCTTATGTAAAGTATATTAACCAAAAGAGGCTTCAAAAGTATGCAATTTAGT 240  
|||||  
Db 300 TGTGCTTATGTAAAGTATATTAACCAAAAGAGGCTTCAAAAGTATGCAATTTAGT 359  
|||||  
QY 241 CAACCTGTGAAGAGCTAATGAATAATCTTGTGCTTTTCAGCTTGACACAGGTTGGAG 300  
|||||  
Db 360 CAACCTGTGAAGAGCTAATGAATAATCTTGTGCTTTTCAGCTTGACACAGGTTGGAG 419  
|||||  
QY 301 TATGCAAAACGCTATATATTTTGCAGAAAAGAAATTAATCTTCTTAACATCTTAAAGAT 360  
|||||  
Db 420 TATGCAAAACGCTATATATTTTGCAGAAAAGAAATTAATCTTCTTAACATCTTAAAGAT 479  
|||||  
QY 361 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGGCCAAAGACCTTCACAGAGT 420  
|||||  
Db 480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGGCCAAAGACCTTCACAGAGT 539  
|||||  
QY 421 GAACCCGAAAATCTCTTCTTGCAGGAACACAGTCTCAGTGTCCCACTCTTAACCTTGA 480  
|||||  
Db 540 GAACCCGAAAATCTCTTCTTGCAGGAACACAGTCTCAGTGTCCCACTCTTAACCTTGA 599  
|||||  
QY 481 ACTGTGAAGACTCTGAGGACAAAGCAGCGGATACAACTTAAAGACGTCTGTACATT 540  
|||||  
Db 600 ACTGTGAAGACTCTGAGGACAAAGCAGCGGATACAACTTAAAGACGTCTGTACATT 659  
|||||  
QY 541 GAATTTGGATCTGATCTTCTTGAAGATACCGTTAATTAAGCACTTATTCAGATGGGGA 600  
|||||  
Db 660 GAATTTGGATCTGATCTTCTTGAAGATACCGTTAATTAAGCACTTATTCAGATGGGGA 719  
|||||  
QY 601 GATCAAGAATTTGTACAAATCACCCCTCAAGAAACAGGATGAATACGTTTGAATTC 660  
|||||  
Db 720 GATCAAGAATTTGTACAAATCACCCCTCAAGAAACAGGATGAATACGTTTGAATTC 779  
|||||  
QY 661 GCAAAAAGGCTGCTTGTGAATTTTCTGACGCGATGTAAACAATCTGAACATCATCAA 720  
|||||  
Db 780 GCAAAAAGGCTGCTTGTGAATTTTCTGACGCGATGTAAACAATCTGAACATCATCAA 839  
|||||  
QY 721 CCCAGTAATTAATGTTGAACACCACTGAGAACCGTGGCCAGAGGCTTCACAGAAAG 780  
|||||  
Db 840 CCCAGTAATTAATGTTGAACACCACTGAGAACCGTGGCCAGAGGCTTCACAGAAAG 899  
|||||  
QY 781 TATCAGGATAGTCTCTTTCAAACTTGCATGTGAGGACATGGCAAAATACTCATGCC 840  
|||||  
Db 900 TATCAGGATAGTCTCTTTCAAACTTGCATGTGAGGACATGGCAAAATACTCATGCC 959  
|||||  
QY 841 AGCTCATTTACAGCATGAGAACACAGCACTTATTAATCACTAAGACAGAAATGTAGAA 900  
|||||  
Db 960 AGCTCATTTACAGCATGAGAACACAGCACTTATTAATCACTAAGACAGAAATGTAGAA 1019  
|||||  
QY 901 AAGGCGAATTTCTGTATTAAGCAAAACGCTGGGCTTGGCAAGGAGCCCAACTTAACAG 960  
|||||  
Db 1020 AAGGCGAATTTCTGTATTAAGCAAAACGCTGGGCTTGGCAAGGAGCCCAACTTAACAG 1079  
|||||  
QY 961 TGGGCTGGAAGTAAAGAACATGTATGATAGGCGGACTCCAGCAGAAAAAAGGTA 1020  
|||||  
Db 1080 TGGGCTGGAAGTAAAGAACATGTATGATAGGCGGACTCCAGCAGAAAAAAGGTA 1139  
|||||  
QY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGAATTAAGCAAGAACTGCCATGC 1080  
|||||  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGAATTAAGCAAGAACTGCCATGC 1199  
|||||  
QY 1081 TCGAGAAATCTTGAAGATACCTGAGAAAGTTCCTTGGATTAACACTAATTAAGCAGCATTCAG 1140  
|||||  
Db 1200 TCGAGAAATCTTGAAGATACCTGAGAAAGTTCCTTGGATTAACACTAATTAAGCAGCATTCAG 1259  
|||||  
QY 1141 AAAGTTAATGAGTGGTTTCCAGAAAGTGAATGAATCTTATAGTTCTGATACATCATGAT 1200  
|||||  
Db 1260 AAAGTTAATGAGTGGTTTCCAGAAAGTGAATGAATCTTATAGTTCTGATACATCATGAT 1319  
|||||  
QY 1201 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGATTTGGAGCTTCTAAATGAGTAGAT 1260  
|||||

```

Db 1320 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGTAATGGACGTTCTAAATGAGTAGAT 1379
Oy 1261 GAATATTCGTGTTCTCAGAGAAAATAGACTTACTGCGCAGTATCCTCATGAGGCTT 1320
Db 1380 GAATATTCGTGTTCTCAGAGAAAATAGACTTACTGCGCAGTATCCTCATGAGGCTT 1439
Oy 1321 ATATGTAAAGTAGAAAAGTAGTCACTCAATCAGTAGAGAGTAATTTGAAGCAAAATA 1380
Db 1440 ATATGTAAAGTAGAAAAGTAGTCACTCAATCAGTAGAGAGTAATTTGAAGCAAAATA 1499
Oy 1381 TTGGGAAAACCTATCGGAAGAGCAACCTCCCACTTAAGCCATGTAAGTGAAT 1440
Db 1500 TTGGGAAAACCTATCGGAAGAGCAACCTCCCACTTAAGCCATGTAAGTGAAT 1559
Oy 1441 CTAAATATAGAGACATTTTCTACTGAGCAGCAGATATATCAAGAGCGTCCCTCACAAT 1500
Db 1560 CTAAATATAGAGACATTTTCTACTGAGCAGCAGATATATCAAGAGCGTCCCTCACAAT 1619
Oy 1501 AAATTAAGCGTAAAGAGAGACCTACATCAGGCGCTTATCTGAGAGATTTTATCAAGAA 1560
Db 1620 AAATTAAGCGTAAAGAGAGACCTACATCAGGCGCTTATCTGAGAGATTTTATCAAGAA 1679
Oy 1561 GGAGATTTGGCAGTTCAAAAGACCTCCTGAATGATATAATCAGAGACTAACCAAGGAG 1620
Db 1680 GGAGATTTGGCAGTTCAAAAGACCTCCTGAATGATATAATCAGAGACTAACCAAGGAG 1739
Oy 1621 CAGAAATGCTCAAGTATGATATTAATTAATAGTGTGATGAGATATAACCAAGAGTGTAT 1680
Db 1740 CAGAAATGCTCAAGTATGATATTAATTAATAGTGTGATGAGATATAACCAAGAGTGTAT 1799
Oy 1681 TCTATTCAGATGAGAAAATCTTAACCAATGATATCAGTCAAGAAAAGAGATCTGCTTC 1740
Db 1800 TCTATTCAGATGAGAAAATCTTAACCAATGATATCAGTCAAGAAAAGAGATCTGCTTC 1859
Oy 1741 AAAAGCAAGCTGAACCTTAAGCAGCAGATATAGCAATATGAGTCAATTAATTAATATC 1800
Db 1860 AAAAGCAAGCTGAACCTTAAGCAGCAGATATAGCAATATGAGTCAATTAATTAATATC 1919
Oy 1801 CACAATTCAAAAGACACCTTAAGAAATAGCTGAGAGAGAGTCTTCTACAGGCAATAT 1860
Db 1920 CACAATTCAAAAGACACCTTAAGAAATAGCTGAGAGAGAGTCTTCTACAGGCAATAT 1979
Oy 1861 CATGGCTTGAACCTGATGATGAGTAAATCTAAGCCCACTTAATTTGTAAGTGAAT 1920
Db 1980 CATGGCTTGAACCTGATGATGAGTAAATCTAAGCCCACTTAATTTGTAAGTGAAT 2039
Oy 1921 ATTGATAGTGTCTTAGCAGTAGAGATTAAGAAAAGAAAAGTACAACCA 1971
Db 2040 ATTGATAGTGTCTTAGCAGTAGAGATTAAGAAAAGAAAAGTACAACCA 2090

RESULT 3
LOCUS B0068830 962 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT_6740238 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5802685
ACCESSION B0068830
VERSION B0068830.1 GI:19897888
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 962)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLCM2037 row: h column: 14  
 High quality sequence stop: 700.  
 Location/Qualifiers

FEATURES  
 source 1. 962  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_xref="IMAGE:5802685"  
 /clone\_lib="NIH\_MGC\_47"  
 /tissue\_type="neuroblastoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAGG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

BASE COUNT 302 a 195 c 235 g 229 t 1 others  
 ORIGIN

Query Match 15.4%; Score 861.4; DB 14; Length 962;  
 Best Local Similarity 96.2%; Pred. No. 1.7e-191;  
 Matches 890; Conservative 0; Mismatches 33; Indels 2; Gaps 1;

```

Oy 4332 TCACAGCAACACACATCGAAGAAAGACGATTAATCTTACAGAAAGTAGTATACC 4391
Db 15 TTCCGACAGAGCAGCTTCAGAAAAGCAGATTAATCTTACAGAAAGTAGTATACC 74
Oy 4392 TATTAACCCAGATCCGAAAGGCTTCTGCTGACAAAGTTGAGTGTGAGATAGTTC 4451
Db 75 TATTAACCCAGATCCGAAAGGCTTCTGCTGACAAAGTTGAGTGTGAGATAGTTC 134
Oy 4452 TACCACTAAAATTAAGAAACAGAGAGTGAAGTGCATCCCTCTTAATGCCATCAT 4511
Db 135 TACCACTAAAATTAAGAAACAGAGAGTGAAGTGCATCCCTCTTAATGCCATCAT 194
Oy 4512 AGATGATAGCTGCTACATGACAGATTTGCTTGGAGTCTTACAGATTAAGATCCATC 4571
Db 195 AGATGATAGCTGCTACATGACAGATTTGCTTGGAGTCTTACAGATTAAGATCCATC 254
Oy 4572 TCAGAGAGAGCTCATTAAGCTTGTGATGTGAGAGAGCAAGCTGGAAGAGTCCGGCC 4631
Db 255 TCAGAGAGAGCTCATTAAGCTTGTGATGTGAGAGAGCAAGCTGGAAGAGTCCGGCC 314
Oy 4632 ACAGATTTGACGAAACATCTTACTTCCCAAGGCAAGATCTAGAGGAGACCCCTTACT 4691
Db 315 ACAGATTTGACGAAACATCTTACTTCCCAAGGCAAGATCTAGAGGAGACCCCTTACT 374
Oy 4692 GGAATGTGAATTCAGCTCTTCTCTGATGAGACCTGAATCTGATCTTGAAGCAGAGC 4751
Db 375 GGAATGTGAATTCAGCTCTTCTCTGATGAGACCTGAATCTGATCTTGAAGCAGAGC 434
Oy 4752 CCCAGAGTACGCTCGTGTGGCAACATACATCTTCAACCTGTGATGGAAGTTCGCCA 4811
Db 435 CCCAGAGTACGCTCGTGTGGCAACATACATCTTCAACCTGTGATGGAAGTTCGCCA 494
Oy 4812 ATTGAAGTTGACAGATTCGCCAGAGTCCAGCTGCTCTATCTACTGATGCTGG 4871
Db 495 ATTGAAGTTGACAGATTCGCCAGAGTCCAGCTGCTCTCTATCTACTGATGCTGG 554
Oy 4872 GTATATATGCAATGGAAGAGTGTGAGCAGAGAGACCAAGTGAAGTGAAGTGAAG 4931
Db 555 TTATATATGCAATGGAAGAGTGTGAGCAGAGAGACCAAGTGAAGTGAAGTGAAG 614
Oy 4932 AAGGCTCAACAAAGAAATGTCATGATGCTGAGTGGCTGAGCCAGAGAAATTAATGCT 4991
Db 615 AAGGCTCAACAAAGAAATGTCATGATGCTGAGTGGCTGAGCCAGAGAAATTAATGCT 674

```

OY	4992	CGTGTACAAGTTTGCCGAAANAACCAATCATTCTTAATCTAATCTAATTACTGAAGAAC	5035
Dd	675	CCTGTACAAGTTTGCCGAAANAACCAATCATTCTTAATCTAATCTAATTACTGAAGAAC	734
OY	5052	TACTCATGTTGGTATGAANAACAGATGCTGCAATTTGTGTGAACGACACTGAATAATTT	5111
Dd	735	TACTCATGTTGGTATGAANAACAGATGCTGCAATTTGTGTGAACGACACTGAATAATTT	794
OY	5112	TCTTAGAATTCGGGAGGAGAAAATGGGTAGTATTTCCTGGGTGACCCTGATATTAA	5171
Dd	795	TCTTAGAATTCGGGAGGAGAAAATGGGTAGTATTTCCTGGGTGACCCTGATATTAA	854
OY	5172	AGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGAGATGG--TCAATGGA	5229
Dd	855	AGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGAGATGGTCAATGGA	914
OY	5230	AGAAACCAACCAAGGCCCAAAGCGAG	5254
Dd	915	AGAAACCAACCAAGGCCCAAAGCGAG	939

[illegible]

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 1089)					
NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .					
National Institutes of Health, Mammalian Gene Collection (MGC)					
Unpublished (1999)					
Contact: Robert Strausberg, Ph.D.					

Email: [cgabps-remail.nih.gov](mailto:cgabps-remail.nih.gov)  
 Tissue Procurement: ATCC/DCTD/DMP  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM12200 row: k column: 19  
 High quality sequence stop: 667.

[illegible]

BASE COUNT	ORIGIN
389 a	212 c 248 g 240 t

Query Match	15.1%	Score 843;	DB 13;	Length 1089;
Best Local Similarity	96.1%;	Pred. No. 3.8e-187;		
Matches 929;	Conservative 0;	Mismatches 30;	Indels 8;	Gaps 6;

1806 TTCAAAAGCACCTAAAAAGAATAGGCTAGAGAGAACTTTCTACCAGGCATATTCATGC 1865

Db  
13 TTCAAAAGCAGCTAAAAAGATTAGGCTGAGGAGAGTCTTCTACGAGCATATTCATGC 72

1866 GCTGACTAGTACGTAGAAATCTAAGCCACCTAATGTACTGATTCGCAATTGA 1925

Db	73	GCCTTAACCTAGTAGTCAGTCAGTAACATCTAAGCCCACTAACTTGTACTAGATTGTGCAATTGGA	132
Qy	1926	TAGTTGTTCTAGCAGCTGAAGAGATAAAGAAAAAGTACACCAAAATCCAGTCAGGCA	1983
Db	133	TAGTTGTTCTAGCAGCTGAAGAGATTAAGCAAAAAAGTACCAACCAAAATCCAGTCAGGCA	192
Qy	1986	CAGCGAAGAAACATCAACATCATGGAAGGTAAAGAACCTGCACATCTGAGGCCAACAGAGTAA	2043
Db	193	CAGCGAAGAAACCTACACACTCATGGAAGGTAAAGAACCTGCACATCTGAGGCCAACAGAGTAA	252
Qy	2046	CAAGCCAAATGAACAGACAGATAAAGAACATGACAGTATCTTCCAGAGCTGAAGTT	2105
Db	253	CAAGCCAAATGAACAGACAGATAAAGAACATGACAGCATACTTCCAGAGCTGAAGTT	312
Qy	2106	AACCAATGCACCTGGTCTTTTCTACTAAGGTCCAATACCGATGAGCACTTAAAGAAATTTGT	2165
Db	313	AACCAATGCACCTGGTCTTTTCTACTAAGGTCCAATACCGATGAGCACTTAAAGAAATTTGT	372
Qy	2166	CAATCCTAGCCTTCCAGAGAGAAAGAAAGAACAGAGAACTAGAACAGTTAAAGTCTCTAA	2223

Qy	2226	TAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGCAACTGAAG	2285
Db	433	TAATCTGAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGCAACTGAAG	492
Qy	2286	ATCTGTAGAGACTAGACAGATTTCATTTGCTACTGCTATTTATGACATCAGAGAA	2345
Db	493	ATCTGTAGAGACTAGACAGATTTCATTTGCTACTGCTATTTATGACATCAGAGAA	552

QY	2346	TATCGGTACTACGGAAGTTAGACACTCTAGGGAAGCAAAAACAGAACAAATAATGTGT	2405
Db	553	TATCGGTACTACGGAAGTTAGACACTCTAGGGAAGCAAAAACAGAACAAATAATGTGT	612
QY	2406	GAGTAGTGTCGACGATTTGAAAACCCCAAGGAGCTAAATTCATGTTGTTCCAAAGATTA	2465
Db	613	GAGTAGTGTCGACGATTTGAAAACCCCAAGGAGCTAAATTCATGTTGTTCCAAAGATTA	672

QY	2466	TAGAAATGACACAGAAGCTTTAAATATCCATTGGACATGAAAGTTAAACCAAGTCGGGA	2522
Db	673	TAGAAATGACACAGAAGCTTTAAATATCCATTGGACATGAAAGTTAAACCAAGTCGGGA	732
QY	2526	AACAAGCATAGAAAT- GGAAGAAATGCAACTTGATGCTCAGATTTTGCAGAAATACATTTCA	2584
Db	733	AACAAGCATAGAAATGGAAGAAATGCAACTTGATGCTCAGATTTTGCAGAAATACATTTCA	792
QY	2585	AGGTTTCAAAGCGCCAGTTCATTTGCTCTGTTTCAAATCCAGGAATCCAGAAGAGAT	2644

Db 793 A-GTTCAAGCGGCAGTCATTTCTCGGTTTCAATCC-GGAATGCAAGAAGGAT 850

[illegible]

RESULT 5	
AU122476	
LOCUS	845 bp
	mRNA
AU122476	linear
	EST 01-AUG-2002

**DEFINITION** A0122476 MAMMA1 Homo sapiens cDNA clone MAMMA1002447 5', mRNA sequence.

ACCESSION	AU122476	.
VERSION	AU122476.1	GI:10937746

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 845)  
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.

TITLE  
JOURNAL  
COMMENT

HRI human cDNA project  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
Location/Qualifiers

FEATURES  
Source

1. 845  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAMMA1002447"  
/clone\_lib="MAMMA1"  
/tissue\_type="mammary gland"  
/note="Vector: PME18SFL3"  
BASE COUNT 279 a 172 c 191 g 200 t 3 others

Query Match 14.5%; Score 811.8; DB 9; Length 845;  
Best Local Similarity 99.1%; Pred. No. 7.5e-180;  
Matches 836; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 3700 GTAACAAATATACCTTCTAGTCTAGTACGATAGCAGCAGCTTCTTCCAGTGTCTCT 3759  
1 GTAACAAATATACCTTCTAGTCTAGTACGATAGCAGCAGCTTCTTCCAGTGTCTCT 60  
DB 3760 AAGAACAAGAGAGAGATTTATATCATTTGAAATAGCTTAATGACTCAGTAAACAG 3819  
61 AAGAACAAGAGAGAGATTTATATCATTTGAAATAGCTTAATGACTCAGTAAACAG 120  
QY 3820 GTAATATTTGGCAAAAGGATCTCAGAAATCATCACCCTTATAGTAAACAATAATCTTCTCT 3879  
121 GTAATATTTGGCAAAAGGATCTCAGAAATCATCACCCTTATAGTAAACAATAATCTTCTCT 180  
QY 3880 AACTTGTCTTCTTCAAGAGAGAGATTTGAAAGAGTCTGCAATATACAAACACCCAG 3939  
181 AACTTGTCTTCTTCAAGAGAGAGATTTGAAAGAGTCTGCAATATACAAACACCCAG 240  
QY 3940 GATCCTTCTTCTTCAAGAGAGATTTGAAAGAGTCTGCAATATACAAACACCCAG 3999  
241 GATCCTTCTTCTTCAAGAGAGATTTGAAAGAGTCTGCAATATACAAACACCCAG 300  
QY 4000 GGTCTGAGTACAGAAATTTGTTTCAATGATGAAAGAAAGAGAAAGAGGCTTGGAGAA 4059  
301 GGTCTGAGTACAGAAATTTGTTTCAATGATGAAAGAAAGAGAAAGAGGCTTGGAGAA 360  
QY 4060 AATATCAAGAGAGAGAGATTTGTTTCAATGATGAAAGAGAGAGAGAGAGAGAGAGAG 4119  
361 AATATCAAGAGAGAGAGATTTGTTTCAATGATGAAAGAGAGAGAGAGAGAGAGAGAG 420  
QY 4120 AATGAAACAG 4179  
421 AATGAAACAG 480  
QY 4180 ACTCAG 4239  
481 ACTCAG 540  
QY 4240 CTAG 4299  
541 CTAG 600

QY 4300 AGTACTCTTCTGCTCCCTTGGAGACCTGCGAAATCCGAAACAAAGACATCAGAAAGCA 4359  
DB 601 AGTACTCTTCTGCTCCCTTGGAGACCTGCGAAATCCGAAACAAAGACATCAGAAAGCA 660  
QY 4360 GTATTAATCTTCAAGAGAGAGAGATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 4419  
DB 661 GTATTAATCTTCAAGAGAGAGAGATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 4420 GCTGACAGAGTGTAGAGTGTCTGAGATAGTCTTACCAAGTAAATTAAGAGAGAGAG 4477  
DB 721 GCTGACAGAGTGTAGAGTGTCTGAGATAGTCTTACCAAGTAAATTAAGAGAGAGAG 780  
QY 4478 TGGAAAGGTCA-TCCCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4536  
DB 781 TGGAAAGGTCAATCCCTCTTCAATTAATCCATCATTTAGATGATAGGNGTNCATGACAGT 840  
QY 4537 TGCT 4540  
DB 841 TGCT 844

RESULT 6  
AUI42729 783 bp mRNA linear EST 05-AUG-2002  
LOCUS AUI42729  
DEFINITION AUI42729 y79AA1 Homo sapiens cDNA clone y79AA1000792 5', mRNA  
ACCESSION AUI42729  
VERSION AUI42729.1 GI:11004250  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.  
TITLE  
JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
Location/Qualifiers

FEATURES  
Source

1. 783  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="y79AA1000792"  
/clone\_lib="y79AA1"  
/cell\_type="retinoblastoma"  
/note="Vector: PME18SFL3"  
BASE COUNT 255 a 146 c 185 g 194 t 3 others

Query Match 13.6%; Score 761.8; DB 9; Length 783;  
Best Local Similarity 99.2%; Pred. No. 4.2e-168;  
Matches 774; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 3460 TTAGATGATGGTGAATAAGAGAGATCTAGTTTGTGAATAAGATTAAGAAAGT 3519  
1 TTAGATGATGGTGAATAAGAGAGATCTAGTTTGTGAATAAGATTAAGAAAGT 60  
DB 3520 TCTGCTTTTGAAG 3579



DB 61 TGTGCTGTTTACGAAAAGCGTCCAGAAAGAGAGCTTACGAGAGCTGCTGACCCCTTTC 120  
QY 3580 ACCCATACACATTTGGCTCAGGTTTACCGAAGAGGGGCCAAGAAATTAGACTCTCAGAA 3639  
DB 121 ACCCATACACATTTGGCTCAGGTTTACCGAAGAGGGGCCAAGAAATTAGACTCTCAGAA 180  
QY 3640 GAGAACTTATCTAGTAGAGATGAAGAGCTTCCCTGCTTCCAACTGTTATTGTTGTA 3699  
DB 181 GGAAGACTTATCTAGTAGAGATGAAGAGCTTCCCTGCTTCCAACTGTTATTGTTGTA 240  
QY 3700 GTAACAAATATACCTTCTAGTCTACTAGCAGTACGACGCTGACCGAGTCTGCT 3759  
DB 241 GTAACAAATATACCTTCTAGTCTACTAGCAGTACGACGCTGACCGAGTCTGCT 300  
QY 3760 AAGAACACAGAGAGATTTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 3819  
DB 301 AAGAACACAGAGAGATTTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 360  
QY 3820 GTAATATTGGCAAGGATCTCAGGAACATCCTTATGAGGAAACAAATGTTCTGCT 3879  
DB 361 GTAATATTGGCAAGGATCTCAGGAACATCCTTATGAGGAAACAAATGTTCTGCT 420  
QY 3880 AGCTGTTTCTTCTCAGTGCAGTGAATGGAAGCTTACGCAATTCACAAACCCAG 3939  
DB 421 AGCTGTTTCTTCTCAGTGCAGTGAATGGAAGCTTACGCAATTCACAAACCCAG 480  
QY 3940 GATCCTTTCTTCTGATGTTCTTCTCACAACAAATGAGGATCAGTCTGAAGCCAGGAGTT 3999  
DB 481 GATCCTTTCTTCTGATGTTCTTCTCACAACAAATGAGGATCAGTCTGAAGCCAGGAGTT 540  
QY 4000 GGTCTGAGTACAGAGATTTGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 4059  
DB 541 GGTCTGAGTACAGAGATTTGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 4060 AATATCAAGAGAGAGAGATTTGATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 4119  
DB 601 AATATCAAGAGAGAGATTTGATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 660  
QY 4120 AGTGAACAGAGGCTCTGTAAGACTGCTCAGAGGCTATCTCTCAGAGTACATTTTAA-C 4178  
DB 661 AGTGAACAGAGGCTCTGTAAGACTGCTCAGAGGCTATCTCTCAGAGTACATTTTAA-C 720  
QY 4179 CACTCAGCAGAGGATPACCATGCAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGA 4238  
DB 721 CACTCAGCAGAGGATPACCATGCAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGA 780  
RESULT 7  
LOCUS BG681276 743 bp mRNA linear EST 01-MAY-2001  
DEFINITION 602627125F1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4751887 5',  
ACCESSION BG681276  
VERSION BG681276.1 GI:13912673  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 743)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-rt@mail.nih.gov](mailto:cgapbs-rt@mail.nih.gov)  
Tissue Procurement: James Cleaver, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM10609 row: a column: 08

FEATURES High quality sequence stop: 741.  
Location/Qualifiers  
source 1..743  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4751887"  
/clone\_lib="NCI\_CGAP\_Skn4"  
/tissue\_type="squamous cell carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1: NotI;  
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 233 a 156 c 182 g 172 t  
ORIGIN  
Query Match 13.2%; Score 740.4; DB 12; Length 743;  
Best Local Similarity 99.9%; Pred. No. 4,4e-163;  
Matches 741; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4605 GGAGCAGACGCTGGAAGAGTGGGCGACAGATTTGACGAAACATCTTACTTGGCAG 4664  
DB 1 GGAGCAGACGCTGGAAGAGTGGGCGACAGATTTGACGAAACATCTTACTTGGCAG 60  
QY 4665 GCAAGATCTAGAGGAGAACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTGATGACCC 4724  
DB 61 GCAAGATCTAGAGGAGAACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTGATGACCC 120  
QY 4725 TGAATCTGATCTTCTGGAAGACAGAGCCCAAGTACGCTGTTGGCAGATACATC 4784  
DB 121 TGAATCTGATCTTCTGGAAGACAGAGCCCAAGTACGCTGTTGGCAGATACATC 180  
QY 4785 TTCAACCTCTGCATGGAAGTGGCCCAATTTGAAGTTGGAGATCTGCCAGAGTCCAGC 4844  
DB 181 TTCAACCTCTGCATGGAAGTGGCCCAATTTGAAGTTGGAGATCTGCCAGAGTCCAGC 240  
QY 4845 TCGCTGCTCATCTACTGATGATGCTGGGTATTAATGCAATGAAGAAAGTGTAGCAGGGA 4904  
DB 241 TCGCTGCTCATCTACTGATGATGCTGGGTATTAATGCAATGAAGAAAGTGTAGCAGGGA 300  
QY 4905 GAAGCCAGATTTGACAGCTTCAACAGAAAGGTTCAACAAAGAAATGTCCATGTTGGTGTG 4964  
DB 301 GAAGCCAGATTTGACAGCTTCAACAGAAAGGTTCAACAAAGAAATGTCCATGTTGGTGTG 360  
QY 4965 TGGCCCTGACCCAGAGAAATTTATGCTGTGTAGCAAGTTGGCAGAAACACCATATCC 5024  
DB 361 TGGCCCTGACCCAGAGAAATTTATGCTGTGTAGCAAGTTGGCAGAAACACCATATCC 420  
QY 5025 TTTAACTAATCTAATTTACTGAAGAGACTACATGTTGTTATGAAGACAGATGCTGAGTT 5084  
DB 421 TTTAACTAATCTAATTTACTGAAGAGACTACATGTTGTTATGAAGACAGATGCTGAGTT 480  
QY 5085 TGTGTGTGACGAGACATGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTTG 5144  
DB 481 TGTGTGTGACGAGACATGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTTG 540  
QY 5145 CTATTTCGGGTGACCCAGTCTATTAAAGAAAGAAATTCGTGATGAGCATGATTTGA 5204  
DB 541 CTATTTCGGGTGACCCAGTCTATTAAAGAAAGAAATTCGTGATGAGCATGATTTGA 600  
QY 5205 AGTCAGAGAGATGTGTCAATGAAGAAACACCAAGAGTCCAAAGCAGAGAGAGATC 5264  
DB 601 AGTCAGAGAGATGTGTCAATGAAGAAAGCAGAGAGTCCAAAGCAGAGAGATC 660  
QY 5265 CCAGAGCAGAAAGATCTTCAGGGGCTTGAAGATCTGTTGCTATGAGGCCCTTACCAACAT 5324  
DB 661 CCAGAGCAGAAAGATCTTCAGGGGCTTGAAGATCTGTTGCTATGAGGCCCTTACCAACAT 720  
QY 5325 GCCCAGACATCAATCGAATGG 5346  
DB 721 GCCCAGACATCAATCGAATGG 742

```

RESULT 8
AUI25312
LOCUS AUI25312 NT2RM4 Homo sapiens cDNA clone NT2RM4001388 5', mRNA
DEFINITION AUI25312 NT2RM4 Homo sapiens cDNA clone NT2RM4001388 5', mRNA
ACCESSION AUI25312
VERSION AUI25312.1 GI:10950028
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Source
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4001388"
/cell_line="teratocarcinoma"
/feature="NT2"
/feature="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 235 a 142 c 163 g 159 t 3 others
ORIGIN
Query Match 12.3%; Score 688; DB 9; Length 702;
Best Local Similarity 9.4%; Pred. No. 9.2e-151;
Matches 699; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 3822 AATATTGCGAAGGATCTCAGGAACATCCTTAGTGAGAAACAAATGTTCTGTAG 3881
Db 1 AATATTGCGAAGGATCTCAGGAACATCCTTAGTGAGAAACAAATGTTCTGTAG 60
QY 3882 CTGTGTTCTCAGAGTGCAGTAATGGAAGACCTGATGCAAAATACAAACCCAGGA 3941
Db 61 CTGTGTTCTCAGAGTGCAGTAATGGAAGACCTGATGCAAAATACAAACCCAGGA 120
QY 3942 TCCTTTCTGATGTTCTTCCAAACAATGAGCATCAGTCTGAAAGCCAGGAGTTGG 4001
Db 121 TCCTTTCTGATGTTCTTCCAAACAATGAGCATCAGTCTGAAAGCCAGGAGTTGG 180
QY 4002 TCTGAGTCAAGAAATGTTGATGATGATGAAGAAAGAGACCGCTTGGAGAAAA 4061
Db 181 TCTGAGTCAAGAAATGTTGATGATGATGAAGAAAGAGACCGCTTGGAGAAAA 240
QY 4062 TAATCAAGAAAGCAAGCATGATGATGATGATGATGATGATGATGATGATGATG 4121
Db 241 TAATCAAGAAAGCAAGCATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 4122 TGAACAGAGCGCTCTGAGAGTCTGAGGCGTATCTCTGAGAGTATTTTAAACAC 4181
Db 301 TGAACAGAGCGCTCTGAGAGTCTGAGGCGTATCTCTGAGAGTATTTTAAACAC 360
QY 4182 TCAGCAGAGGATACCATGCAACATACCTGATTAAGCTCCAGCAGGAATGCTGAACT 4241

```

```

Db 361 TCAGCAGAGGATACCATGCAACATACCTGATTAAGCTCCAGCAGGAATGCTGAACT 420
QY 4242 AGAAGCTGTGTAGAAACATGAGGAGCCAGCTTCTAAGCTTCCCTTCATCATAG 4301
Db 421 AGAAGCTGTGTAGAAACATGAGGAGCCAGCTTCTAAGCTTCCCTTCATCATAG 480
QY 4302 TGACTTCTGCTTGTGAGGAGCCAGCTTCTAAGCTTCCCTTCATCATAG 4361
Db 481 TGACTTCTGCTTGTGAGGAGCCAGCTTCTAAGCTTCCCTTCATCATAG 540
QY 4362 ATTAACTTCAAGAAAGTAGTGAATACCTTAAAGCCAGAAATCCAGAGCCCTTTCTGC 4421
Db 541 ATTAACTTCAAGAAAGTAGTGAATACCTTAAAGCCAGAAATCCAGAGCCCTTTCTGC 600
QY 4422 TCAGCAGGTTGAGGCTGTGAGATAGTCTTCCAGTAAATAAGAACAGAGGTGA 4481
Db 601 TCAGCAGGTTGAGGCTGTGAGATAGTCTTCCAGTAAATAAGAACAGAGGTGA 660
QY 4482 AAGGTCATCCCTTCTAATGCCCATCATATGATGATAGTGG 4524
Db 661 AANGNCAT-CCCTTCTAATGCCCATCATATGATGATAGNG 702
RESULT 9
BF508987/c
LOCUS BF508987 739 bp mRNA linear EST 06-DEC-2000
DEFINITION UI-H-B14-aot-b-06-0-UI-s1 NCI-GAP_Sub8 Homo sapiens cDNA clone
IMAGE:3085787 3', mRNA sequence.
ACCESSION BF508987
VERSION BF508987.1 GI:11592285
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 739)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE NCI-CGAP Sub8 library derived from NCI-CGAP Sub8.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgap@bbs-rmail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LINL at:
www.bio.linnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=yes.
FEATURES
Source
1..739
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3085787"
/cell_line="NCI-GAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/feature="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI-CGAP_Sub8
is a subtracted library derived from NCI-CGAP Sub8. The
NCI-CGAP_Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI-CGAP_Sub8 was used
as a tracer in a subtractive hybridization with a driver
comprising: a pool of clones from NCI-CGAP_Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population), a pool of clones from NCI-CGAP_Sub4
(IMAGE clone ids 2723592-2729326; 25% of the driver
population), NCI-CGAP_Sub6 (pool AIF-AU; IMAGE ids
2728969-2733190; 25% of the driver population), and
NCI-CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550
; 25% of the driver population). Subtraction was

```

performed as previously described [Bonafide, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6: 791-806.

"TAG-SBQ-found"

BASE COUNT 173 a 141 c 128 g 295 t 2 others

Query Match 12.3%; Score 686.6; DB 12; Length 739;  
Best Local Similarity 99.1%; Pred. No. 2e-150; Mismatches 6; Gaps 0;  
Matches 688; Conservative 0; Indels 0; Gaps 0;

1267 TCTGTTCTTCAGAGAAATAGACTTCTGCGCAGTATCTCATGAGCCTTAAATATGT 1326  
1268 TCTGTTCTTCAGAGAAATAGACTTCTGCGCAGTATCTCATGAGCCTTAAATATGT 1326  
1327 AAAAGTGAAGAGTTTCTTCCAAATCAGTAGAGATTAATTTGAGACAAATATTGGG 1386  
638 AAAAGTGAAGAGTTTCTTCCAAATCAGTAGAGATTAATTTGAGACAAATATTGGG 579  
1387 AAAAGTGAAGAGTTTCTTCCAAATCAGTAGAGATTAATTTGAGACAAATATTGGG 1446  
578 AAAAGTGAAGAGTTTCTTCCAAATCAGTAGAGATTAATTTGAGACAAATATTGGG 519  
1447 ATAGAGCATTTTGTACTGAGCCACAGATTAATTAACAAGAGCCTCCCTCAATAATTA 1506  
518 ATAGAGCATTTTGTACTGAGCCACAGATTAATTAACAAGAGCCTCCCTCAATAATTA 459  
1507 AAGCGTAAAGAGAGCTTACATCAGGCTTCTCATCTGAGAGATTTTATCAAGAAAGCAGAT 1566  
458 AAGCGTAAAGAGAGCTTACATCAGGCTTCTCATCTGAGAGATTTTATCAAGAAAGCAGAT 399  
1567 TTGGGAGTCAAAAGAGCTCTGGAATGATTAATCAAGAGAACTTAACCAAGGAGCAGAT 1626  
398 TTGGGAGTCAAAAGAGCTCTGGAATGATTAATCAAGAGAACTTAACCAAGGAGCAGAT 339  
1627 GGTCAAGTATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 1686  
338 GGTCAAGTATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 279  
1687 CAGATGAGAAATCTTAACCAATGATGATGATGATGATGATGATGATGATGATGAT 1746  
278 CAGATGAGAAATCTTAACCAATGATGATGATGATGATGATGATGATGATGATGAT 219  
1747 AAGCTGATCTATTAAGAGAGTATTAATGATGATGATGATGATGATGATGATGAT 1806  
218 AAGCTGATCTATTAAGAGAGTATTAATGATGATGATGATGATGATGATGATGAT 159  
1807 TCAAAAGAGCTTAAAGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1866  
158 TCAAAAGAGCTTAAAGAGTATGATGATGATGATGATGATGATGATGATGATGAT 99  
1867 CTTGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1926  
98 CTTGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 39  
1927 AGTTGTTCTGAGAGTGAAGATTAAGAAAGAAAA 1961  
38 AGTTGTTCTGAGAGTGAAGATTAAGAAAGAAAA 4

RESULT 10  
BM800251 987 bp mRNA linear EST 05-MAR-2002  
LOCUS BM800251 AGENCOURT 6415998 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5531286  
DEFINITION 5' mRNA sequence.  
ACCESSION BM800251  
VERSION BM800251.1 GI:19117074  
KEYWORDS EST.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 987)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLM12212 row: 1 column: 07  
High quality sequence stop: 692.

## FEATURES

## source

Location/Qualifiers

1..987

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5531286"

/clone\_lib="NIH\_MGC-71"

/issue\_type="telomysarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.1 kb."

BASE COUNT 342 a 181 c 222 g 238 t 4 others

Query Match 12.3%; Score 686.4; DB 14; Length 987;  
Best Local Similarity 85.8%; Pred. No. 2.5e-150;  
Matches 874; Conservative 0; Mismatches 14; Indels 131; Gaps 5;

279 TCACTTGACACAGCTTGGAGTATGCAACACCTTAATTTGCAAAAGAGAAATTA 338  
1 TAGACTTGACACAGCTTGGAGTATGCAACACCTTAATTTGCAAAAGAGAAATTA 60  
339 CTCTCTGGAACATCTTAAGATGATGATGATGATGATGATGATGATGATGATGAT 398  
61 CTCTCTGGAACATCTTAAGATGATGATGATGATGATGATGATGATGATGATGAT 120  
399 TGCAGAAAGAGCTTCAAGAGTGAACCGGAAATCTCTCTGCGAGAAACAGTCTAG 458  
121 TGCAGAAAGAGCTTCAAGAGTGAACCGGAAATCTCTCTCTGCGAGTCTAG 177  
459 TGTCCACACTCTTAACCTTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 518  
178 TGTCCACACTCTTAACCTTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 237  
519 TCAAAAGAGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 578  
238 TCAAAAGAGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 264  
579 GGCACCTTATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 638  
265 -----GGCTGCTTGTGAATTTCTGAGACGATGT 264  
639 GATGAATACGTTGGATTTCGAAAAAGAGCTGTTGTGAATTTCTGAGACGATGT 698  
265 -----GGCTGCTTGTGAATTTCTGAGACGATGT 294  
699 AACAAATACTGAACATCATCAACCCAGTATATGATTTGAACACACAGAGAGGTCG 758  
295 AACAAATACTGAACATCATCAACCCAGTATATGATTTGAACACACAGAGAGGTCG 354  
759 AGCTGAGAGCATCCAGAAAGATATCAGAGTGTCTGTTCAAACTTCATGAGAGCC 818  
355 AGCTGAGAGCATCCAGAAAGATATCAGAGTGTCTGTTCAAACTTCATGAGAGCC 414  
819 ATGTGGCAAAATACTCATGCTCAGCTCATTTAAGCATGGAACAGCAGTATTACTAC 878  
415 ATGTGGCAAAATACTCATGCTCAGCTCATTTAAGCATGGAACAGCAGTATTACTAC 474

```

QY 879 TAAAGCAGAAATGAAATGTAGAAAAGCGTAATCTTGTATATAAGCAACAGCTGGCTT 938
D 475 TAAAGCAGAAATGAAATGTAGAAAAGCGTAATCTTGTATATAAGCAACAGCTGGCTT 934
QY 939 AGCAAGGAGCAACATACAGATGGGCTGGAAGTAAGAAATGTAATGATAGCGGAC 998
D 535 AGCAAGGAGCAACATACAGATGGGCTGGAAGTAAGAAATGTAATGATAGCGGAC 994
QY 999 TCCCGACAGAAAAAAGTATGATCTGATCCCTGCTGTGAGAGAAAAGATG 1058
D 595 TCCCGACAGAAAAAAGTATGATCTGATCCCTGCTGTGAGAGAAAAGATG 654
QY 1059 GAATAGAGCAAAATGCGATGCTCAGAGAACTCTAGAGATGTAAGATGTTCTTGAT 1118
D 655 GAATAGAGCAAAATGCGATGCTCAGAGAACTCTAGAGATGTAAGATGTTCTTGAT 714
QY 1119 AACCTAATATGACATTCAGAAAGTAAATGAGAGGTTTCCAGAGATGATGATGTT 1178
D 715 AACCTAATATGACATTCAGAAAGTAAATGAGAGGTTTCCAGAGATGATGATGTT 774
QY 1179 AGGTTCGATGACTCAGATGATGGGAGTCTGAATCAATGCCAAAGTACGT-GATGTAT 1237
D 775 AGGTTCGATGACTCAGATGATGGGAGTCTGAATCAATGCCAAAGTACGTGATGTAT 834
QY 1238 TGGACGTT-CTTAATGAGGTAGATG---AATATTTGTTGTTCTTCAGAGAAAATAGACTT 1292
D 835 TGGACGTTCTTAATGAGGTAGATGATGATGTTGTTGTTCTTCAGAGAAAATAGACTT 893

```

```

RESULT 11
LOCUS B0422380 899 bp mRNA linear EST 23-MAY-2002
DEFINITION AGNCOURT_7802085 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042052
5', mRNA sequence.
B0422380
VERSION B0422380.1 GI:21117695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 899)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13280 row: n column: 05
High quality sequence stop: 597.
Location/Qualifiers
1. 899

```

```

FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6042052"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 291 a 182 c 192 g 230 t 4 others
ORIGIN
Query Match 11.9% Score 666.6 DB 14 Length 899;

```

```

Best Local Similarity 97.5%; Pred. No. 1,1e-145;
Matches 740; Conservative 0; Mismatches 11; Indels 8; Gaps 6;

```

```

QY 1 ATGAGTTTATCTGCTCTCCGTTGAGAGATACAAAATGCTAATGATGCACAGAAA 60
D 100 ATGAGTTTATCTGCTCTCCGTTGAGAGATACAAAATGCTAATGATGCACAGAAA 159
QY 61 ATCTTAGAGTGTCCATCTGCTGTGAGTGTATCAAGAACTCTCTCCAAAGTGTAC 120
D 160 ATCTTAGAGTGTCCATCTGCTGTGAGTGTATCAAGAACTCTCTCCAAAGTGTAC 219
QY 121 CACATATTTTGCATTTTTCATGCTGAACTCTTCAACCAAGAAAGAGGCTTCACAG 180
D 220 CACATATTTTGCATTTTTCATGCTGAACTCTTCAACCAAGAAAGAGGCTTCACAG 279
QY 181 TGTCTTATATGTAATATGATATATCAAAAGAGAGCTTCAAGAAAGATGATAGT 240
D 280 TGTCTTATATGTAATATGATATATCAAAAGAGAGCTTCAAGAAAGATGATAGT 339
QY 241 CAACCTGTTGAAGAGCTATGTAATAATCATTTTGTCTTCAAGTGTACAGAGTTTGAG 300
D 340 CAACCTGTTGAAGAGCTATGTAATAATCATTTTGTCTTCAAGTGTACAGAGTTTGAG 399
QY 301 TATGCAACAGCTATATTTTTCAGAAAAGAAATTAATCTCTGAAACATCTTAAAGAT 360
D 400 TATGCAACAGCTATATTTTTCAGAAAAGAAATTAATCTCTGAAACATCTTAAAGAT 459
QY 361 GAAGTTTCTATCATCAAAAGATGAGGCTACAGAAACGCTGCCAAAGACCTTACAGAGT 420
D 460 GAAGTTTCTATCATCAAAAGATGAGGCTACAGAAACGCTGCCAAAGACCTTACAGAGT 519
QY 421 GAACCCGAAATCCTCTCTGAGAAACAGCTCAGTGTCAACCTCTTAAACCTTGA 480
D 520 GAACCCGAAATCCTCTCTGAGAAACAGCTCAGTGTCAACCTCTTAAACCTTGA 576
QY 481 ACTGTGAGAACTCTGAGCAAAAGCAGCGGATTAACACTCAAAAGACGTCTGTACATT 540
D 577 ACTGTGAGAACTCTGAGCAAAAGCAGCGGATTAACACTCAAAAGACGTCTGTACATT 636
QY 541 GAATGGGATGATGATTTCTTGAAGATACCGTTAATAAGCAATTTTTCAGGTGGA 600
D 637 GAATGGGATGATGATTTCTTGAAGATACCGTTAATAAGCAATTTTTCAGGTGGA 656
QY 601 GATCAAGATTTGTAACAATCACCCCTCAAGAACCA-GGATGAAATCAGTTTGATTC 659
D 697 GATCAAGATTTGTAACAATCACCCCTCAAGAACCA-GGATGAAATCAGTTTGATTC 756
QY 660 TGCAGAAAAGGCTGCTGTTGAATTTTCTGAGAGCGATGTACAA-ATACTGAACATC-A 716
D 757 TGCAGAAAAGGCTGCTGTTGAATTTTCTGAGAGCGATGTACAA-ATACTGAACATC-A 816
QY 717 TCACCCAGTATATATGA-TTTGAACAGCACTGAGAGC 754
D 817 TCACCCAGTATATATGA-TTTTGAACAGCACTGAGAGC 855

```

```

RESULT 12
LOCUS BF791668 900 bp mRNA linear EST 12-JAN-2001
DEFINITION 602251822F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4344317 5'
mRNA sequence.
B0791668
VERSION BF791668.1 GI:12096722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1AM962 row: k column: 06  
 High quality sequence stop: 678.  
 Location/Qualifiers

## FEATURES

source

1. 900  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:434317"  
 /clone\_1lb="NIH\_MGC\_84"  
 /tissue\_type="adrenal cortex carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 283 a 176 c 212 g 229 t

ORIGIN

Query Match 11.8%; Score 657; DB 12; Length 900;

Best Local Similarity 96.5%; Pred. No. 2e-143; Matches 736; Conservative 0; Mismatches 20; Indels 7; Gaps 6;

QY 3356 CTGTTAATACAGATTCTCTCATATCTGATTTTCAGATTAAGACACCTATGGAA 3415  
 Db 1 CTGTATATACAGATTCTCTCATATCTGATTTTCAGATTAAGACACCTATGGAA 60  
 QY 3416 GTAGTCATGCATCTCAGGTTGTTCTGAGACACCTGATGACTGTATGATGATGAA 3475  
 Db 61 GTAGTCATGCATCTCAGGTTGTTCTGAGACACCTGATGACTGTATGATGATGAA 120  
 QY 3476 TAAAGAGATGATGATGTTGCTGAAATGACATTAAGAAAGTTCGCTGTTTTCGA 3535  
 Db 121 TAAAGAGATGATGATGTTGCTGAAATGACATTAAGAAAGTTCGCTGTTTTCGA 180  
 QY 3536 AAAGCGTCAGAAAG 3595  
 Db 181 AAAGCGTCAG 240  
 QY 3596 CTGAGGTTACCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3655  
 Db 241 CTGAGGTTACCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 QY 3656 AGGATGAAGAGCTCCCTGCTCCAAACACTGTTATTTGTTAAAGTAACATATACCT 3715  
 Db 301 AGGATGAAGAGCTCCCTGCTCCAAACACTGTTATTTGTTAAAGTAACATATACCT 359  
 QY 3716 CTCAGTCTACTAGGATGACAGCGTTGCTACCGAGTGTCTGTCTAAGACAGAGAGA 3775  
 Db 360 CTCAGTCTACTAGGATGACAGCGTTGCTACCGAGTGTCTGTCTAAGACAGAGAGA 419  
 QY 3776 ATTATTTATCTGAGAGATAGCTTAATGACTGCTAATCAACAGATATATTTGGCAA 3835  
 Db 420 ATTATTTATCTGAGAGATAGCTTAATGACTGCTAATCAACAGATATATTTGGCAA 478  
 QY 3836 CATCTCGAAGACATCCTAGTGAAGAGAAACAAATGCTGCTAGTCTTTCTTCAC 3895  
 Db 479 CATCTCGAAGACATCCTAGTGAAGAGAAACAAATGCTGCTAGTCTTTCTTCAC 538  
 QY 3896 AGTGCATGATTTGAGAGACTGCTGCAATATCAACACACAGAGATCTTTCTGATTG 3955  
 Db 539 AGTGCATGATTTGAGAGACTGCTGCAATATCAACACACAGAGATCTTTCTGATTG 598  
 QY 3956 GTTCTTCAACAAGATGAGCATCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4015  
 Db 599 GTTCTTCAACAAGATGAGCATCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656

QY 4016 AATGCTTTACATGATGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4075  
 Db 657 AA-TGCTTTACATGATGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714  
 QY 4076 AAAGCATGATTCAAACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4118  
 Db 715 CAAGCATGAG-TTCACACTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756

## RESULT 13

BG178466

LOCUS

DEFINITION 602330271F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4431568 5', mRNA sequence.

ACCESSION

BG178466

VERSION

BG178466.1

KEYWORDS

EST, human.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 921)

AUTHORS

NIH-MGC http://imgc.ncbi.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1AM10186 row: n column: 17  
 High quality sequence stop: 631.  
 Location/Qualifiers

FEATURES

source

1. 921  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4431568"  
 /clone\_1lb="NIH\_MGC\_91"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT

340 a 173 c 205 g 203 t

ORIGIN

Query Match 11.1%; Score 619.2; DB 12; Length 921;

Best Local Similarity 90.0%; Pred. No. 1.5e-134; Matches 734; Conservative 0; Mismatches 68; Indels 14; Gaps 6;

QY 2623 CCAGGAATCCAGAGAGAGATGTCACATTTCTGCCCCACTGTGGCTCTTAAGAAA 2682  
 Db 11 CCAGGAATCCAGAGAGAGATGTCACATTTCTGCCCCACTGTGGCTCTTAAGAAA 70  
 QY 2683 CAAAGTCCAAAGCACTTTTGAATGCAACAAAGAGAGAGAGAGAGAGAGAG 2742  
 Db 71 CAAAGTCCAAAGCACTTTTGAATGCAACAAAGAGAGAGAGAGAGAGAGAGAG 130  
 QY 2743 TCTAATATCAAGCTGTACAGACAGTAAATATCACTGACGCTTCTGTGTTGTCAG 2802  
 Db 131 TCTAATATCAAGCTGTACAGACAGTAAATATCACTGACGCTTCTGTGTTGTCAG 190  
 QY 2803 AAAGATTAAGCAGTTGATTAATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTA 2862  
 Db 191 AAAGATTAAGCAGTTGATTAATGCCAAATGTAGTATCAAGAGAGGCTCTAGG-TTTGTCTA 249

2863 TCATCTCAGTTCAGAGGCAACGAAGTGCATCTACTTCCAAATAAATGAGCTTTTA 2922  
|||||  
250 TCATCTCAGTTCAGAGGCAACGAAGTGCATCTACTTCCAAATAAATGAGCTTTTA 309  
|||||  
2923 CAAACCCATTCGTATCCACACCTTTTCCATCAAGTCATTTGTTAAACTAATCT 2982  
|||||  
310 CAAACCCATTCGTATCCACACCTTTTCCATCAAGTCATTTGTTAAACTAATCT 369  
|||||  
2983 AAGAAATCTCTGCTAGAGGAAACTTTGAGAACATTCATCTCAGCTGAAGAGAAATG 3042  
|||||  
370 AAGAAATCTCTGCTAGAGGAAACTTTGAGAACATTCATCTCAGCTGAAGAGAAATG 429  
|||||  
3043 GGAATATGAGACATTCAGTCAAGTACAGTACAGTCAATTAAGCCGTAATACATTTAGAGAAAT 3102  
|||||  
430 GGAATATGAGACATTCAGTCAAGTACAGTACAGTCAATTAAGCCGTAATACATTTAGAGAAAT 489  
|||||  
3103 GTTTTAAAGAGCCGCTCAAGCATATTTAATGAAGTAGGTTCCAGTACTAATGAAGT 3162  
|||||  
490 GTTTTAAAGAGCCGCTCAAGCATATTTAATGAAGTAGGTTCCAGTACTAATGAAGT 549  
|||||  
3163 GCCTCAGTATTAATGAATAGGTTCCAGTACAGTCAATTAAGCCGTAATACATTTAGAGAAAT 3222  
|||||  
550 GCCTCAGTATTAATGAATAGGTTCCAGTACAGTCAATTAAGCCGTAATACATTTAGAGAAAT 609  
|||||  
3223 AACAGAGGCGCAAAATGATGCTATGCTTAGATTAAGGGGTTTGCACCTGAGG- 3277  
|||||  
610 AACAGAGGCGG- AAATGATGATGCTTAGATTAAGGGGTTTGCACCAACCGTGAGGTC 668  
|||||  
3278 -TCTTAAACAAAGTCTTCTGGAAGTAAATGTAACCATTCGTAATTAAGGCAAGAA 3336  
|||||  
669 TATTAACCAAGGCTTCTGGAAGTAAATGTAACCATTCGTAATTAAGGCAAGAA 726  
|||||  
3337 TATGAAGAATGATGCTGAGCTGTTAATACAGATTTCTCTCAGTCTGATTCAGATTAAC 3396  
|||||  
727 ATTTGAGAGTGTGTC- ACCGGAACAGATTTTCCCCCATCT- -GATCCGATAC 782  
|||||  
3397 TTGAAGACGCTATGGAAGTGTGTCATCTCAG 3432  
|||||  
783 TTAAACCGCATGGAAGACGCAACTAGTTGAG 818  
|||||

RESULT 14  
LOCUS BQ215100 878 bp mRNA linear EST 02-MAY-2002  
DEFINITION AGENCOURT\_7591049 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:605516  
5', mRNA sequence.  
ACCESSION BQ215100  
VERSION BQ215100.1 GI:20396500  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 878)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: ATCC/DC/DP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM1341 row: 0 column: 21  
High quality sequence stop: 669.  
Location/Qualifiers  
1..878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:605516"

FEATURES  
Source

/clone\_lib="NIH\_MGC\_72"  
/issue\_type="melanotic melanoma"  
/lab\_host="PH10B (phage-resistance)"  
/note="Organ: skin; Vector: pcwv-sport6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 281 a 180 c 190 g 227 t  
ORIGIN

Query Match 11.0%; Score 616.2; DB 14; Length 878;  
Best Local Similarity 98.7%; Pred. No. 7.6e-134;  
Matches 664; Conservative 0; Mismatches 3; Indels 6; Gaps 4;

1 ATGATTTATTCGCTTCGCTGGAAGTAAAGTAAATGCTAATGCTATGACAGAA 60  
|||||  
103 ATGATTTATTCGCTTCGCTGGAAGTAAAGTAAATGCTAATGCTATGACAGAA 162  
|||||  
61 ATCTAGAGTGTCCATCTGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTATGAC 120  
|||||  
163 ATCTAGAGTGTCCATCTGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTATGAC 222  
|||||  
121 CACATATTTTGCATATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCT 180  
|||||  
223 CACATATTTTGCATATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCT 282  
|||||  
181 TGTCTTTATGTAAGTAAATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTAATG 240  
|||||  
283 TGTCTTTATGTAAGTAAATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTAATG 342  
|||||  
241 CACATTTTGCATATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTAATG 300  
|||||  
343 CACATTTTGCATATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCTAATG 402  
|||||  
301 TATGCAACAGCTATTAATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCT 360  
|||||  
403 TATGCAACAGCTATTAATTTTGCATGCTGGAAGTAAAGTAAAGTAAATGCTAATGCT 462  
|||||  
361 GAAGTTTCTATCATCCAAAGTATGGAAGTAAAGTAAAGTAAATGCTAATGCTAATGCT 420  
|||||  
463 GAAGTTTCTATCATCCAAAGTATGGAAGTAAAGTAAAGTAAATGCTAATGCTAATGCT 522  
|||||  
421 GAACCCGAAATCTCTTCTGCAAGAACAGTCTGATGTCACATCTCTAACCTTGA 480  
|||||  
523 GAACCCGAAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 579  
|||||  
481 ACTGTGAGACGCTGAGGCAAGGAGGATGACAACTCAAAAGAGCTGCTTACATTT 540  
|||||  
580 ACTGTGAGACGCTGAGGCAAGGAGGATGACAACTCAAAAGAGCTGCTTACATTT 639  
|||||  
541 GAATGGAGATCTGATTTCTTGAAGATACCGTTAATAGGCAACTTATTTGAGT- GTGG 599  
|||||  
640 GAATGGAGATCTGATTTCTTGAAGATACCGTTAATAGGCAACTTATTTGAGT- GTGG 699  
|||||  
600 AGATCAAGATTTGTTACAAATCACCCTCAAGAGCA- GGAATGAATTCAG- TTGGAT 657  
|||||  
700 AGATCAAGATTTGTTACAAATCACCCTCAAGAGCA- GGAATGAATTCAG- TTGGAT 759  
|||||  
658 TCTGCAAAAAAGG 670  
|||||  
760 TCTGCAAAAAAG 772  
|||||

RESULT 15  
LOCUS AI992040/c 630 bp mRNA linear EST 08-SEP-1999  
DEFINITION ws44c12.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2500054 3  
similar to gb:U14680 BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN  
(HUMAN);, mRNA sequence.  
ACCESSION AI992040  
VERSION AI992040.1 GI:5838945  
KEYWORDS EST.  
SOURCE human.







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 12:34:02 ; Search time 13911.7 Seconds

(Without alignments)  
11692.029 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_120\_5708

Perfect score: 5589  
Sequence: 1 ATGATTATTCCTCTCTCG.....AGATCCCCACAGCCACTAC 5589Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

```
1:  gb_da:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_ov:*
5:  gb_ov:*
6:  gb_ov:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_mu:*
22: em_mu:*
23: em_mu:*
24: em_mu:*
25: em_mu:*
26: em_mu:*
27: em_mu:*
28: em_mu:*
29: em_mu:*
30: em_mu:*
31: em_mu:*
32: em_mu:*
33: em_mu:*
34: em_mu:*
35: em_mu:*
36: em_mu:*
37: em_mu:*
38: em_mu:*
39: em_mu:*
40: em_mu:*
41: em_mu:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5587.4	100.0	5711	6 AR007334	AR007334 Sequence
2	5587.4	100.0	5711	6 AR112810	AR112810 Sequence
3	5585.8	99.9	5711	6 AR033056	AR033056 Sequence
4	5585.8	99.9	5711	9 HSU14680	U14680 Homo sapien
5	5585.8	99.9	5712	6 AR070223	AR070223 Sequence
6	5585.8	99.9	5712	6 AR118507	AR118507 Sequence
7	5585.8	99.9	5712	6 AR125601	AR125601 Sequence
8	5585.8	99.9	5712	6 AR184044	AR184044 Sequence
9	5585.8	99.9	5914	6 AR004673	AR004673 Sequence
10	5585.8	99.9	5914	6 AR008159	AR008159 Sequence
11	5585.8	99.9	5914	6 AR136942	AR136942 Sequence
12	5585.8	99.9	5914	6 AR16943	AR16943 Sequence
13	5585.8	99.9	5914	6 AR0938	AR0938 Sequence
14	5585.8	99.9	5914	6 AR048660	AR048660 Sequence
15	5584.2	99.9	5711	6 AR048666	AR048666 Sequence
16	5584.2	99.9	5711	6 AR0795	AR0795 Sequence
17	5584.2	99.9	5711	6 AR0795	AR0795 Sequence
18	5584.2	99.9	5711	6 AR0795	AR0795 Sequence
19	5582.6	99.9	5711	6 AR007335	AR007335 Sequence
20	5582.6	99.9	5711	6 AR112809	AR112809 Sequence
21	5581	99.9	5711	6 AR007333	AR007333 Sequence
22	5581	99.9	5711	6 AR112808	AR112808 Sequence
23	5581	99.9	5711	6 AR048668	AR048668 Sequence
24	5574.8	99.7	5712	6 AR0803	AR0803 Sequence
25	5574.8	99.7	5710	6 AR048662	AR048662 Sequence
26	5573.8	99.7	5710	6 AR048662	AR048662 Sequence
27	5573.8	99.7	5710	6 AR048662	AR048662 Sequence
28	5571.8	99.7	5709	6 AR048658	AR048658 Sequence
29	5571.8	99.7	5709	6 AR048663	AR048663 Sequence
30	5571.8	99.7	5709	6 AR048664	AR048664 Sequence
31	5571.8	99.7	5709	6 AR048665	AR048665 Sequence
32	5571.8	99.7	5709	6 AR0793	AR0793 Sequence
33	5571.8	99.7	5709	6 AR0798	AR0798 Sequence
34	5571.8	99.7	5709	6 AR0799	AR0799 Sequence
35	5571.8	99.7	5709	6 AR0800	AR0800 Sequence
36	5567.8	99.6	5707	6 AR048667	AR048667 Sequence
37	5567.8	99.6	5707	6 AR0802	AR0802 Sequence
38	5535.8	99.0	5709	6 AR281859	AR281859 Sequence
39	5531.8	99.0	5689	6 AR048659	AR048659 Sequence
40	5531.8	99.0	5689	6 AR0794	AR0794 Sequence
41	5526.6	98.9	5589	9 AF207822	AF207822 Pan trogl
42	5507.8	98.5	5693	9 AF005068	AF005068 Homo sapi
43	5505.6	98.5	5770	6 AR048661	AR048661 Sequence
44	5505.6	98.5	5770	6 AR048661	AR048661 Sequence
45	5467.8	97.8	5656	6 AR048657	AR048657 Sequence

## ALIGNMENTS

```
RESULT 1
AR007334 LOCUS 5711 bp DNA Linear PAT 04-DEC-1998
DEFINITION Sequence 3 from patent US 5750400.
ACCESSION AR007334
VERSION AR007334.1 GI:3966818
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5711)
AUTHORS Murphy,P.D., Allen,A.C., Alvares,C.P., Critz,B.S., Olson,S.J.,
Schelter,D.B. and Zeng,B.
TITLE Coding sequences of the human BRCA1 gene
JOURNAL Patent: US 5750400-A 3 12-MAY-1998;
```

FEATURES Location/Qualifiers  
source 1..5711  
BASE COUNT 1956 a 1098 c 1274 g 1383 t  
ORIGIN  
Query Match 100.0%; Score 5587.4; DB 6; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Caps 0;  
QY 1 AAGGATTTTCGCTTGGCGCTGAAGAGTCAAAATGCTATATGCTATGACGAAA 60  
Db 120 AAGGATTTATTCGCTTGGCGCTGAAGAGTCAAAATGCTATATGCTATGACGAAA 179  
QY 61 ATCTTAGAGTGTCCATCTGTCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 120  
Db 180 ATCTTAGAGTGTCCATCTGTCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 239  
QY 121 CACATATTTTGCATAATTTTGCATGCTGAACCTTCTCAACCCAGAAAGGCCCTTCACAG 180  
Db 240 CACATATTTTGCATAATTTTGCATGCTGAACCTTCTCAACCCAGAAAGGCCCTTCACAG 299  
QY 181 TGTCCCTTTATGTAAGATATATATACCAAAAGAGGCTACAGAAAGTACGAAATTTAGT 240  
Db 300 TGTCCCTTTATGTAAGATATATATACCAAAAGAGGCTACAGAAAGTACGAAATTTAGT 359  
QY 241 CAACCTTGTGAAGAGCTATTGAATAATCATTTGTGCTTTCAGCTGACACAGGTTTGAG 300  
Db 360 CAACCTTGTGAAGAGCTATTGAATAATCATTTGTGCTTTCAGCTGACACAGGTTTGAG 419  
QY 301 TATGCAACAGCTATATATTTTGCAAAAAGAAATTAACCTCTCTCAACATCTTAAAGAT 360  
Db 420 TATGCAACAGCTATATATTTTGCAAAAAGAAATTAACCTCTCTCAACATCTTAAAGAT 479  
QY 361 GAAGTTTTCATATCCAAAGTATGAGCTACAGAAACCGTCCAAAAGCTTCTACAGAGT 420  
Db 480 GAAGTTTTCATATCCAAAGTATGAGCTACAGAAACCGTCCAAAAGCTTCTACAGAGT 539  
QY 421 GAACCCGAAATCCCTCTCTTGCAGGAAACCAAGTCTCAGTCCCACTCTTAACCTTGA 480  
Db 540 GAACCCGAAATCCCTCTCTTGCAGGAAACCAAGTCTCAGTCCCACTCTTAACCTTGA 599  
QY 481 ACTGTGAGAGCTCTGAGCAAAAGCAGCGGATACAACTCTCAAAAGCCTCTCTCAAT 540  
Db 600 ACTGTGAGAGCTCTGAGCAAAAGCAGCGGATACAACTCTCAAAAGCCTCTCTCAAT 659  
QY 541 GAATGGGATCGATTTCTCTGAGATACCGTTAATGAAGCAACTATGCAAGTGGGA 600  
Db 660 GAATGGGATCGATTTCTCTGAGATACCGTTAATGAAGCAACTATGCAAGTGGGA 719  
QY 601 GATCAGAATTTGTTCAAAATCACCCCTCAAGGAACCGAGGATGAATCAGTTTGATTC 660  
Db 720 GATCAGAATTTGTTCAAAATCACCCCTCAAGGAACCGAGGATGAATCAGTTTGATTC 779  
QY 661 GCAAAAAAGGCTGCTTGTGAATTTTCTGAGCGGATGAACAAATCTGAACATCTCA 720  
Db 780 GCAAAAAAGGCTGCTTGTGAATTTTCTGAGCGGATGAACAAATCTGAACATCTCA 839  
QY 721 CCCAGTAAATGATTTGAAACACCACTGAGAGCGGCTGAGAGGATCCAGAAAG 780  
Db 840 CCCAGTAAATGATTTGAAACACCACTGAGAGCGGCTGAGAGGATCCAGAAAG 899  
QY 781 TATCAGGGTAGTCTGTTCAAACTTGCATGTGAGCCATGTGGCAACAATCTACTATGCC 840  
Db 900 TATCAGGGTAGTCTGTTCAAACTTGCATGTGAGCCATGTGGCAACAATCTACTATGCC 959  
QY 841 AGCTATTTACGAGTACGAGACAGCAGTTTATTACTACTAAGACAGATGAATGTAGAA 900  
Db 960 AGCTATTTACGAGTACGAGACAGCAGTTTATTACTACTAAGACAGATGAATGTAGAA 1019  
QY 901 AAGGCTGAATCTGTAATTAAGCAAAACAGCCTGGCTTAGCAAGAGCCAAATATACAA 960  
Db 1020 AAGGCTGAATCTGTAATTAAGCAAAACAGCCTGGCTTAGCAAGAGCCAAATATACAA 1079

QY 961 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGAGCTCCGACAGAAAAAAGGTA 1020  
Db 1080 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGAGCTCCGACAGAAAAAAGGTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAATTAAGCAGAACTGCCATGC 1080  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAATTAAGCAGAACTGCCATGC 1199  
QY 1081 TCAGAGATTCCTAGAGATCTAGAGATGCTTCTGTGATTAACCTTAATTAAGCAGATTCAG 1140  
Db 1200 TCAGAGATTCCTAGAGATCTAGAGATGCTTCTGTGATTAACCTTAATTAAGCAGATTCAG 1259  
QY 1141 AAAAGTTAATGAGTGTGTTTCCAGAGTGAAGTAACTGTTAGGTTCTGATGACTCAATGAT 1200  
Db 1260 AAAAGTTAATGAGTGTGTTTCCAGAGTGAAGTAACTGTTAGGTTCTGATGACTCAATGAT 1319  
QY 1201 GGGGAGTCTGAATCAATATCCAAAGTAGCTGATGATTGAGAGCTTCTAATGAGGTAGAT 1260  
Db 1320 GGGGAGTCTGAATCAATATCCAAAGTAGCTGATGATTGAGAGCTTCTAATGAGGTAGAT 1379  
QY 1261 GAATATTCGTGTTCTTCAAGAAATTAAGTACTAGGCGAGTAACTCATGAGGCTTTA 1320  
Db 1380 GAATATTCGTGTTCTTCAAGAAATTAAGTACTAGGCGAGTAACTCATGAGGCTTTA 1439  
QY 1321 ATATGTAAAGTGAAGAGTTCATCTCAAAATCAGTAGAGAGTAAATTTGAAGACAAATA 1380  
Db 1440 ATATGTAAAGTGAAGAGTTCATCTCAAAATCAGTAGAGAGTAAATTTGAAGACAAATA 1499  
QY 1381 TTTGGGAAAACTATCGGAGAAAGCAGCCCTCCCACTTAAGCAGTAACTGAAT 1440  
Db 1500 TTTGGGAAAACTATCGGAGAAAGCAGCCCTCCCACTTAAGCAGTAACTGAAT 1559  
QY 1441 CTATATATAGAGCATTTGTTACTAGGCGACAGATTAATTAAGAGCGCCCTCAACAAT 1500  
Db 1560 CTATATATAGAGCATTTTCTACTAGGCGACAGATTAATTAAGAGCGCCCTCAACAAT 1619  
QY 1501 AAATTAAGCGTAAAGAGAGACCTACATCAGGCTTCAATCCAGAGATTTTATCAAGAA 1560  
Db 1620 AAATTAAGCGTAAAGAGAGACCTACATCAGGCTTCAATCCAGAGATTTTATCAAGAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAAGCTCTCGAATGATTAATCAAGGAACTTAACCAAGCAG 1620  
Db 1680 GCAGATTTGGCAGTTCAAAAAGCTCTCGAATGATTAATCAAGGAACTTAACCAAGCAG 1739  
QY 1621 CAGAAATGTCAGAGTGAATTAATTAATGTTGTCATGAGAAATTAACCAAGAGTAT 1680  
Db 1740 CAGAAATGTCAGAGTGAATTAATTAATGTTGTCATGAGAAATTAACCAAGAGTAT 1799  
QY 1681 TCTATTCAGAAATGAGAAAAATCCTAACCCATAGAAATCACTCGAAAAAAGATTCGCTTTC 1740  
Db 1800 TCTATTCAGAAATGAGAAAAATCCTAACCCATAGAAATCACTCGAAAAAAGATTCGCTTTC 1859  
QY 1741 AAAAGCAAAAGCTGAACCTATTAAGACAGCAGTATAGCAATATGAACTCGAATTAATATC 1800  
Db 1860 AAAAGCAAAAGCTGAACCTATTAAGACAGCAGTATATACCAATATGAACTCGAATTAATATC 1919  
QY 1801 CACAATTCAAAAAGCACTTAAAAAGAAATAGGCGTGAAGAGAGTCTTACACGAGCATAT 1860  
Db 1920 CACAATTCAAAAAGCACTTAAAAAGAAATAGGCGTGAAGAGAGTCTTACACGAGCATAT 1979  
QY 1861 CATGGCCTTGAACCTAGTATGCTAGTAATTAAGCCCACTTAATTTACTGATTCGAA 1920  
Db 1980 CATGGCCTTGAACCTAGTATGCTAGTAATTAAGCCCACTTAATTTACTGATTCGAA 2039  
QY 1921 ATTGATAGTGTGTTCTAGCAGTGAAGAGATTAAGAAAAAGTCAACCAATATCCAGTC 1980  
Db 2040 ATTGATAGTGTGTTCTAGCAGTGAAGAGATTAAGAAAAAGTCAACCAATATCCAGTC 2099  
QY 1981 AGGCACAGAGAACTCAACTCATGAGAGTAAAGAACTGCACTGAGAGCCCAAG 2040  
Db 2100 AGGCACAGAGAACTCAACTCATGAGAGTAAAGAACTGCACTGAGAGCCCAAG 2159

2041 AGTAACAGCCAAATGACAGACAGTAAGTAAGACATGACAGTACTTCCAGAGCTG 2100  
2160 AGTAACAGCCAAATGACAGACAGTAAGTAAGACATGACAGTACTTCCAGAGCTG 2219  
2101 AAGTTAACAATGACAGCTGTTCTTTACTAAGTGTCAATACAGTGAACCTTAAGAA 2160  
2220 AAGTTAACAATGACAGCTGTTCTTTACTAAGTGTCAATACAGTGAACCTTAAGAA 2279  
2161 TTGTCAATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
2280 TTGTCAATCTGAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2339  
2221 TCTAATATGCTGAAG 2280  
2340 TCTAATATGCTGAAG 2399  
2281 GAAAGATCTGTAG 2340  
2400 GAAAGATCTGTAG 2459  
2341 GAAAGATCTGTAG 2400  
2460 GAAAGATCTGTAG 2519  
2401 TGTGTAGTCACTGTGACAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
2520 TGTGTAGTCACTGTGACAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2579  
2461 GATTAATGAGATGACAG 2520  
2580 GATTAATGAGATGACAG 2639  
2521 GGGAGAGAGAGATGAG 2580  
2640 GGGAGAGAGAGATGAG 2699  
2581 TTCAAGGTTTCAAG 2640  
2700 TTCAAGGTTTCAAG 2759  
2641 GAATGTGACAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700  
2760 GAATGTGACAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2819  
2701 TTGTAATGTAAG 2760  
2820 TTGTAATGTAAG 2879  
2761 CAGAGAGATTAATATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820  
2880 CAGAGAGATTAATATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2939  
2821 AATGCCAATATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880  
2940 AATGCCAATATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2999  
2881 AAGGAAGTGAAGTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940  
3000 AAGGAAGTGAAGTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3059  
2941 CCACCACTTTTCCATCAAGTCAATTTGTTAAACTAATGTAAGAGAGAGAGAGAG 3000  
3060 CCACCACTTTTCCATCAAGTCAATTTGTTAAACTAATGTAAGAGAGAGAGAGAG 3119  
3001 GAAAGATCTGTAG 3060  
3120 GAAAGATCTGTAG 3179  
3061 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
3180 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3239  
3121 TCAGCAATATTAATGAG 3180

3240 TCAGCAATATTAATGAG 3299  
3181 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
3300 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3359  
3241 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
3360 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3419  
3301 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
3420 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3479  
3361 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
3480 AATAGAGTGAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3539  
3421 CATGATCTGAG 3480  
3540 CATGATCTGAG 3599  
3481 GAGAGATTAATGAG 3540  
3600 GAGAGATTAATGAG 3659  
3541 GTCAG 3600  
3660 GTCAG 3719  
3601 GGTACCGAG 3660  
3720 GGTACCGAG 3779  
3661 GAGAGATTAATGAG 3720  
3780 GAGAGATTAATGAG 3839  
3721 TCTAG 3780  
3840 TCTAG 3899  
3781 TCTAG 3840  
3900 TCTAG 3959  
3841 CAGAGAGATTAATGAG 3900  
3960 CAGAGAGATTAATGAG 4019  
3901 AATGAGATTAATGAG 3960  
4020 AATGAGATTAATGAG 4079  
3961 TCCAAACAATGAG 4020  
4080 TCCAAACAATGAG 4139  
4021 GTTTCAGATGAG 4080  
4140 GTTTCAGATGAG 4199  
4081 ATGAGATTAATGAG 4140  
4200 ATGAGATTAATGAG 4259  
4141 GACTGCTAG 4200  
4260 GACTGCTAG 4319  
4201 CAACATTAATTAATGAG 4260

Db 4320 CAACATACTGATTAAGCTCCAGCAGGAAATGGCTGAACCTAGAACTGTGTAGAACG 4379  
QY 4261 CATGGAGCGACGCTTCTTAACAGCTACCCCTTCATCATTAAGACGCTCTGCGCTTGAG 4320  
Db 4380 CATGGAGCGACGCTTCTTAACAGCTACCCCTTCATCATTAAGACGCTCTGCGCTTGAG 4439  
QY 4321 GACCTGCGAAATCCAGAACCAACACATCAGAAAAAGCAGTATTAATTCACAGAAAGT 4380  
Db 4440 GACCTGCGAAATCCAGAACCAACACATCAGAAAAAGCAGTATTAATTCACAGAAAGT 4499  
QY 4381 AGTGAATACCTATTAAGCAGAAATCCAGAGGCTTCTGCTGACAGATTTGAGGTCT 4440  
Db 4500 AGTGAATACCTATTAAGCAGAAATCCAGAGGCTTCTGCTGACAGATTTGAGGTCT 4559  
QY 4441 GCAGATAGTCTACCAAGTAAATTAAGAACCCAGAGAGTGAAGGTCATCCCTCTTAA 4500  
Db 4560 GCAGATAGTCTACCAAGTAAATTAAGAACCCAGAGAGTGAAGGTCATCCCTCTTAA 4619  
QY 4501 TGCCCATATTAGATGATAGGTGTACATGCACAGTTGCTGGAGTCTTGAGATAGA 4560  
Db 4620 TGCCCATATTAGATGATAGGTGTACATGCACAGTTGCTGGAGTCTTGAGATAGA 4679  
QY 4561 AACTACCCATCTCAAGAGAGCTCATTAAGCTTGTGATGTGAGAGCAGACAGCTGGA 4620  
Db 4680 AACTACCCATCTCAAGAGAGCTCATTAAGCTTGTGATGTGAGAGCAGACAGCTGGA 4739  
QY 4621 GAGTCTGGGCGACAGATTTTGACGGAACATCTTACTTGCCCAAGCAAGATCTAGAGGA 4680  
Db 4740 GAGTCTGGGCGACAGATTTTGACGGAACATCTTACTTGCCCAAGCAAGATCTAGAGGA 4799  
QY 4681 ACCCTTAACCTGGAATCTGGAATCAGCCTCTTCTGATGACCCCTGAATCTGATCTT 4740  
Db 4800 ACCCTTAACCTGGAATCTGGAATCAGCCTCTTCTGATGACCCCTGAATCTGATCTT 4859  
QY 4741 GAAGCAGAGGCGCCAGAGTCAAGCTGCTGTGTCACATACCATCTTCAACCTCTGAT 4800  
Db 4860 GAAGCAGAGGCGCCAGAGTCAAGCTGCTGTGTCACATACCATCTTCAACCTCTGAT 4919  
QY 4801 AAAGTCCCAATTTGAAGTTGACAAATCTGCCAGAGTCCAGCTGCTGCTCATCTACT 4860  
Db 4920 AAAGTCCCAATTTGAAGTTGACAAATCTGCCAGAGTCCAGCTGCTGCTCATCTACT 4979  
QY 4861 GATACCTGCTGGGTATATGCAATGAGAAAGTGTGACAGGAGGAGAGCCAGATTTGACA 4920  
Db 4980 GATACCTGCTGGGTATATGCAATGAGAAAGTGTGACAGGAGGAGAGCCAGATTTGACA 5039  
QY 4921 GCTTCAACAGAAAGGCTCAACAAAAGATGTCATGGTGGTGGCTGACCCACAGA 4980  
Db 5040 GCTTCAACAGAAAGGCTCAACAAAAGATGTCATGGTGGTGGCTGACCCACAGA 5099  
QY 4981 GAATTTATGCTGCTGTACAAAGTTGCCAGAAACACACATCACTTAACTAATCTAAT 5040  
Db 5100 GAATTTATGCTGCTGTACAAAGTTGCCAGAAACACACATCACTTAACTAATCTAAT 5159  
QY 5041 ACTGAAGAGCTACTCATGTTGTTATGAAAACAGATGCTGATTTGTGTGACAGGACA 5100  
Db 5160 ACTGAAGAGCTACTCATGTTGTTATGAAAACAGATGCTGATTTGTGTGACAGGACA 5219  
QY 5101 CTGAATAATTTTCTAGGAATGGGAGGAGAAATGCGTAGTATTTCTGGGTGACC 5160  
Db 5220 CTGAATAATTTTCTAGGAATGGGAGGAGAAATGCGTAGTATTTCTGGGTGACC 5279  
QY 5161 CAGCTCTTTTAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG 5220  
Db 5280 CAGCTCTTTTAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG 5339  
QY 5221 GTCAATGAGAAACCAACCAAGGTCCAAAGCGAGCAGAGAGATCCAGAGAGAGAAATC 5280  
Db 5340 GTCAATGAGAAACCAACCAAGGTCCAAAGCGAGCAGAGAGATCCAGAGAGAGAAATC 5399  
QY 5281 TTCAAGGGGCTAGAAATCTGTTGCTATGAGGCTTACCAACATGCCCACAGATCACTG 5340  
Db 5400 TTCAAGGGGCTAGAAATCTGTTGCTATGAGGCTTACCAACATGCCCACAGATCACTG 5459

QY 5341 GAATGAGAGTACAGCTGTGTGCTCTGTGAGTGAAGAGGCTTTCATCATCACCTT 5400  
Db 5460 GAATGAGAGTACAGCTGTGTGCTCTGTGAGTGAAGAGGCTTTCATCATCACCTT 5519  
QY 5401 GGCACAGGTGTCACCAATGTTGGTTGTGACAGCCAGATGCTGAGACAGAAAGG 5460  
Db 5520 GGCACAGGTGTCACCAATGTTGGTTGTGACAGCCAGATGCTGAGACAGAAAGG 5579  
QY 5461 TTCCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCAGAGTGGTTGGAC 5520  
Db 5580 TTCCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCAGAGTGGTTGGAC 5639  
QY 5521 AGTGTAGACCTCTACAGTGGCAGAGGCTGAGACACTCTGATACCCAGATCCCCAC 5580  
Db 5640 AGTGTAGACCTCTACAGTGGCAGAGGCTGAGACACTCTGATACCCAGATCCCCAC 5699  
QY 5581 AGCCACTAC 5589  
Db 5700 AGCCACTAC 5708

RESULT 2  
AR112810  
LOCUS AR112810 5711 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 5 from patent US 6130322.  
ACCESSION AR112810  
VERSION AR112810.1 GI:14092710  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5711)  
AUTHORS Murphy, P.D., Allen, A.C.P., Alvares, C.P., Critz, B.S., Olson, S.J.,  
Thurber, D., and Jeng, B.  
TITLE Coding sequences of the human BRCA1 gene  
JOURNAL Patent: US 6130322-A 5 10-OCT-2000;  
FEATURES  
source 1..5711  
BASE COUNT 1956 a 1098 c 1274 g 1383 t  
ORIGIN

Query Match 100.0%; Score 5587.4; DB 6; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGATTTTATCTGCTCTTCCGCTTGAAGAAAGTACAAAATGTCATTAAATGCTATGAGAAA 60  
Db 120 ATGATTTTATCTGCTCTTCCGCTTGAAGAAAGTACAAAATGTCATTAAATGCTATGAGAAA 179  
QY 61 ATCTTAGAGTGTCCCATCTGTCTGAGTGTGATCAGAGAACCTGCTCCACAAAGTGTGAC 120  
Db 180 ATCTTAGAGTGTCCCATCTGTCTGAGTGTGATCAGAGAACCTGCTCCACAAAGTGTGAC 239  
QY 121 CACATATTTTGCATATTTTGCATGCTGAAACTTCTCAACAGAAAGAGGCTTCCACAG 180  
Db 240 CACATATTTTGCATATTTTGCATGCTGAAACTTCTCAACAGAAAGAGGCTTCCACAG 239  
QY 181 TGTCCCTTTATGTAAGATGATTAACCAAAAGAGGCTTCAAGAAAGTACAGAGATTTAGT 240  
Db 300 TGTCCCTTTATGTAAGATGATTAACCAAAAGAGGCTTCAAGAAAGTACAGAGATTTAGT 359  
QY 241 CACCTGTGTAAGAGCTATTGAANAATCATTTGTGCTTTTACGCTGTGACAGAGTGGAG 300  
Db 360 CACCTGTGTAAGAGCTATTGAANAATCATTTGTGCTTTTACGCTGTGACAGAGTGGAG 419  
QY 301 TATGCAAAACAGCTTAATTTTGAAGAAAGAAATTAATCTGCTGAACATCTAAAGAT 360  
Db 420 TATGCAAAACAGCTTAATTTTGAAGAAAGAAATTAATCTGCTGAACATCTAAAGAT 479  
QY 361 GAACTTCTATCATCAAAAGTATGAGCTAGAGAAACCGTCCAAAGACTTCTACAGAGT 420

D 480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAGACTTTACAGAGT 539  
Q 421 GAACCCGAATTCCTCTGCGAAGAACAGCTCAGTGTCCACTCTTACCTTGA 480  
D 540 GAACCCGAATTCCTCTGCGAAGAACAGCTCAGTGTCCACTCTTACCTTGA 599  
Q 481 ACTGTGAGAACTCTGAGAACAGCGGATACAACTCTCAAAAGAGCTGTCTCAT 540  
D 600 ACTGTGAGAACTCTGAGAACAGCGGATACAACTCTCAAAAGAGCTGTCTCAT 659  
Q 541 GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATAGGCAACTTATGCTGGA 600  
D 660 GAATTTGGATCTGATTTCTTCTGAGATACCGTTAATAGGCAACTTATGCTGGA 719  
Q 601 GATCAAGATTTGTACAAATCACCCCTCAGAAACAGGATGAATCACTTGGATCT 660  
D 720 GATCAAGATTTGTACAAATCACCCCTCAGAAACAGGATGAATCACTTGGATCT 779  
Q 661 GCAAAAAGGCTCTGTGATTTTCTGAGACGATGTAACTCAATCTGAACATCA 720  
D 780 GCAAAAAGGCTCTGTGATTTTCTGAGACGATGTAACTCAATCTGAACATCA 839  
Q 721 CCCAGTATATATATTTGAACACCACTGAGACGCTGAGAGGATCCAGAAAG 780  
D 840 CCCAGTATATATATTTGAACACCACTGAGACGCTGAGAGGATCCAGAAAG 899  
Q 781 TATCAGGATGTTCTGTTCAAACTGATGAGACGATGAGCAATCTGATCC 840  
D 900 TATCAGGATGTTCTGTTCAAACTGATGAGACGATGAGCAATCTGATCC 959  
Q 841 AGCTCATTCAGCAGTGAACACAGCTTTATCTACTAAAGACAGATGATGAG 900  
D 960 AGCTCATTCAGCAGTGAACACAGCTTTATCTACTAAAGACAGATGATGAG 1019  
Q 901 AAGCGTATTTCTGTATATAAGCAACAGCTGCTGAGCAAGAGCAATACAG 960  
D 1020 AAGCGTATTTCTGTATATAAGCAACAGCTGCTGAGCAAGAGCAATACAG 1079  
Q 961 TGGGCTGGAAGTGAAGAACATGATGATGAGCGGACCTCCAGCAAGAAAGG 1020  
D 1080 TGGGCTGGAAGTGAAGAACATGATGATGAGCGGACCTCCAGCAAGAAAGG 1139  
Q 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATAGCAGAACTGC 1080  
D 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATAGCAGAACTGC 1199  
Q 1081 TCGAGAAATCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1140  
D 1200 TCGAGAAATCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 1259  
Q 1141 AAGGTAATGAGTGTCTTCCAGAGTGAACGTGATGATGATGATGATGATGAT 1200  
D 1260 AAGGTAATGAGTGTCTTCCAGAGTGAACGTGATGATGATGATGATGATGAT 1319  
Q 1201 GGGGAGTCTGAATCAATGCCAAAGTAGTGTGATGATGATGATGATGATGAT 1260  
D 1320 GGGGAGTCTGAATCAATGCCAAAGTAGTGTGATGATGATGATGATGATGAT 1379  
Q 1261 GAATATTTGCTGCTTCAAGAAATAGATGATGATGATGATGATGATGATGAT 1320  
D 1380 GAATATTTGCTGCTTCAAGAAATAGATGATGATGATGATGATGATGATGAT 1439  
Q 1321 ATATGTAAGTGAAGAGTGTCTCCTCAATCAGTGAAGTGAATGTAAGCAAA 1380  
D 1440 ATATGTAAGTGAAGAGTGTCTCCTCAATCAGTGAAGTGAATGTAAGCAAA 1499  
Q 1381 TTTGGGAAACCTATCGAAGAGGCAAGCTCCCACTTAAGCCATGATGAA 1440  
D 1500 TTTGGGAAACCTATCGAAGAGGCAAGCTCCCACTTAAGCCATGATGAA 1559  
Q 1441 CTAAATATAGAGCATTTGTACTGAGCCAGATATACAGAGGCTCCCTCA 1500  
D 1560 CTAAATATAGAGCATTTGTACTGAGCCAGATATACAGAGGCTCCCTCA 1619

Q 1501 AAATTAAGGCTAAAGAGAGACCTACATCAGGCTTCATCCTGAGATTTATCA 1560  
D 1620 AAATTAAGGCTAAAGAGAGACCTACATCAGGCTTCATCCTGAGATTTATCA 1679  
Q 1561 GCAGATTTGGCAGTCTCAAAAGACTCTGAAATGATAATCAGGGAACCAAC 1620  
D 1680 GCAGATTTGGCAGTCTCAAAAGACTCTGAAATGATAATCAGGGAACCAAC 1739  
Q 1621 CAGATGCTCAAGTGAATATTTACTAATAGTGTATGATGATGATGATGAT 1680  
D 1740 CAGATGCTCAAGTGAATATTTACTAATAGTGTATGATGATGATGATGAT 1799  
Q 1681 TCTATTCAGATGAGAAATCTTAACCCCAATGATCCTGCAAAAGAGTGTCT 1740  
D 1800 TCTATTCAGATGAGAAATCTTAACCCCAATGATCCTGCAAAAGAGTGTCT 1859  
Q 1741 AAACGAAGCTGACCTATTAACAGCAGTATTAAGCAATGAGACTCGAATTA 1800  
D 1860 AAACGAAGCTGACCTATTAACAGCAGTATTAAGCAATGAGACTCGAATTA 1919  
Q 1801 CACATTTCAAAAGCACCCTAAAGAAATAGGCTGAGAGAGTCTTACAGGCA 1860  
D 1920 CACATTTCAAAAGCACCCTAAAGAAATAGGCTGAGAGAGTCTTACAGGCA 1979  
Q 1861 CATGCGCTGAACTAGTACGATGATGATGATGATGATGATGATGATGAT 1920  
D 1980 CATGCGCTGAACTAGTACGATGATGATGATGATGATGATGATGATGAT 2039  
Q 1921 ATTGATAGTGTCTGTAGCAGTGAAGATTAAGCAAAAGAAAGTAAACCA 1980  
D 2040 ATTGATAGTGTCTGTAGCAGTGAAGATTAAGCAAAAGAAAGTAAACCA 2099  
Q 1981 AGGCACAGCAAACTACACATCTATGAGAGTAAAGAACCTGACGCAAGAG 2040  
D 2100 AGGCACAGCAAACTACACATCTATGAGAGTAAAGAACCTGACGCAAGAG 2159  
Q 2041 AGTAACAGCAAACTACACATCTATGAGAGTAAAGAACCTGACGCAAGAG 2100  
D 2160 AGTAACAGCAAACTACACATCTATGAGAGTAAAGAACCTGACGCAAGAG 2219  
Q 2101 AAGTAACAATGACCTGCTTTTCTTACTAAGTGTCAATACGATGAACCTAA 2160  
D 2220 AAGTAACAATGACCTGCTTTTCTTACTAAGTGTCAATACGATGAACCTAA 2279  
Q 2161 TTTGTCAATCTGCTTCCAGAGAGAAAGAAAGAACTAAGACGTTAAAGT 2220  
D 2280 TTTGTCAATCTGCTTCCAGAGAGAAAGAAAGAACTAAGACGTTAAAGT 2339  
Q 2221 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTT 2280  
D 2340 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAGGTTT 2399  
Q 2281 GAAAGATCTGTAGAGAGTATTTCAATGATGATGATGATGATGATGATGAT 2340  
D 2400 GAAAGATCTGTAGAGAGTATTTCAATGATGATGATGATGATGATGATGAT 2459  
Q 2341 GAAAGTATCTGTTACTGGAAGTATGATGATGATGATGATGATGATGATGAT 2400  
D 2460 GAAAGTATCTGTTACTGGAAGTATGATGATGATGATGATGATGATGATGAT 2519  
Q 2401 TGTGTAGTCAAGTGTGAGAGTGTGAAGACCCCAAGGACTAATGATGTTT 2460  
D 2520 TGTGTAGTCAAGTGTGAGAGTGTGAAGACCCCAAGGACTAATGATGTTT 2579  
Q 2461 GATATAGAAATGACAGAGGCTTTAAGTATCATTTGGACATGAAGTTAAC 2520  
D 2580 GATATAGAAATGACAGAGGCTTTAAGTATCATTTGGACATGAAGTTAAC 2639  
Q 2521 CGGGAACAAGCTATGAATGGAAGAGTGAAGCTGATGATGATGATGATGAT 2580  
D 2640 CGGGAACAAGCTATGAATGGAAGAGTGAAGCTGATGATGATGATGATGAT 2699

2581 TTCAGGTTTCACAAAGCCGACATCTTTGCTGCTTTTCAATCCAGGAAATGCAGAGAG 2640  
|||||  
2700 TTCAGGTTTCACAAAGCCGACATCTTTGCTGCTTTTCAATCCAGGAAATGCAGAGAG 2759  
2641 GAATGTCACAACTCTCTGCCACCTCTGGTCTCTTAAGAAACAAAGTCCAAAGTCAC 2700  
|||||  
2760 GAATGTCACAACTCTCTGCCACCTCTGGTCTCTTAAGAAACAAAGTCCAAAGTCAC 2819  
2701 TTTGATGTCACAAAGGAGAAATCAAGGAAAGATGATCTATATCAAGCCCTGA 2760  
|||||  
2820 TTTGATGTCACAAAGGAGAAATCAAGGAAAGATGATCTATATCAAGCCCTGA 2879  
2761 CAGACAGTTAATATCACTCAGGCTTCTGCTGTTGTCAGAAAGATAGCCAGTTGAT 2820  
|||||  
2880 CAGACAGTTAATATCACTCAGGCTTCTGCTGTTGTCAGAAAGATAGCCAGTTGAT 2939  
2821 AATGCAATGTAGTATCAAGGAGGCTAGGTTTGTCTATCATCTCAGTTCAGAGGC 2880  
|||||  
2940 AATGCAATGTAGTATCAAGGAGGCTAGGTTTGTCTATCATCTCAGTTCAGAGGC 2999  
2881 AACGAACCTGACTCTATCTCCCAATTAACATGATGATTTTACAAACCCATATCTATA 2940  
|||||  
3000 AACGAACCTGACTCTATCTCCCAATTAACATGATGATTTTACAAACCCATATCTATA 3059  
2941 CCACCACTTTTCCCATCAAGTCATTTGTTAAAACTTAATGTAGAAAAATCTGCTAGAG 3000  
|||||  
3060 CCACCACTTTTCCCATCAAGTCATTTGTTAAAACTTAATGTAGAAAAATCTGCTAGAG 3119  
3001 GAAACCTTTGAGGACATTCATGTGACCGGAAAGGAAATGGGAATGGAACATTTCCA 3060  
|||||  
3120 GAAACCTTTGAGGACATTCATGTGACCGGAAAGGAAATGGGAATGGAACATTTCCA 3179  
3061 AGTAGAGTGCAGCAATTTAGCCGTAATTAACATTTAGGAAATGTTTAAAGAACCCAGC 3120  
|||||  
3180 AGTAGAGTGCAGCAATTTAGCCGTAATTAACATTTAGGAAATGTTTAAAGAACCCAGC 3239  
3121 TCACCAATATTAAAGAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAAGAA 3180  
|||||  
3240 TCACCAATATTAAAGAGTACGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAAGAA 3299  
3181 ATAGATTCAGTATGAAACAACTTCAGAGGAACTGTGAGAAACAGAGGGCCAAATATG 3240  
|||||  
3300 ATAGATTCAGTATGAAACAACTTCAGAGGAACTGTGAGAAACAGAGGGCCAAATATG 3359  
3241 AATGCTATGCTAGATTAGGGGTTTTCGAACCTGAGGCTATTAACAAAGTCTTCCCTGA 3300  
|||||  
3360 AATGCTATGCTAGATTAGGGGTTTTCGAACCTGAGGCTATTAACAAAGTCTTCCCTGA 3419  
3301 AGTAATTTGAGCATCTCGAATAATAAAGCAAGAAATATGAGAGAGTACTGAGACTGT 3360  
|||||  
3420 AGTAATTTGAGCATCTCGAATAATAAAGCAAGAAATATGAGAGAGTACTGAGACTGT 3479  
3361 AATACAGATTTCTCCATATGATGATTTAGATTAACAGGCTATGGAAGTATGAT 3420  
|||||  
3480 AATACAGATTTCTCCATATGATGATTTAGATTAACAGGCTATGGAAGTATGAT 3539  
3421 CATGATCTCAGGTTGTTCTAGAGACCTGATGACCTGTAGATGATGATGATGATGAT 3480  
|||||  
3540 CATGATCTCAGGTTGTTCTAGAGACCTGATGACCTGTAGATGATGATGATGATGAT 3599  
3481 GAAGTACTAGTTTGTCTGAAAATGACATTAAGGAAAGTTCTGCTTTTTCGCAAAAGC 3540  
|||||  
3600 GAAGTACTAGTTTGTCTGAAAATGACATTAAGGAAAGTTCTGCTTTTTCGCAAAAGC 3659  
3541 GTCCAGAAAGAGAGCTTAGAGAGGCTAGCCCTTACCCATACCAATTTGGCTGAG 3600  
|||||  
3660 GTCCAGAAAGAGAGCTTAGAGAGGCTAGCCCTTACCCATACCAATTTGGCTGAG 3719  
3601 GGTTCACCAAGAGGGCCAAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTGGAT 3660  
|||||  
3720 GGTTCACCAAGAGGGCCAAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTGGAT 3779  
3661 GAAGAGCTTCCCTGCTCCAAACACTTGTATTGTTAAAGTAAACATATACCTTCTGAG 3720

|||||  
3780 GAAGAGCTTCCCTGCTCCAAACACTTGTATTGTTAAAGTAAACAAATATACCTTCTAG 3839  
3721 TCTACTAGGATATGACACCGTTGCTACGAGTGTCTGTCTTAACACACAGAGAGAAATTA 3780  
|||||  
3840 TCTACTAGGATATGACACCGTTGCTACGAGTGTCTGTCTTAACACACAGAGAGAAATTA 3899  
3781 TTATCATTTGAGAGATAGCTTAAATGACTGACATACAGATGATTAATTTGGCAAGGATCT 3840  
|||||  
3900 TTATCATTTGAGAGATAGCTTAAATGACTGACATACAGATGATTAATTTGGCAAGGATCT 3959  
3841 CAGGAACATCACCCTTAGTGAAGGAAACAAATGTTCTGCTAGCTTTTCTTACAGTGC 3900  
|||||  
3960 CAGGAACATCACCCTTAGTGAAGGAAACAAATGTTCTGCTAGCTTTTCTTACAGTGC 4019  
3901 AGTGAATTTGGAAGACTGACCTGCAATATCAACACCCAGGATCTTCTGATGTTGCT 3960  
|||||  
4020 AGTGAATTTGGAAGACTGACCTGCAATATCAACACCCAGGATCTTCTGATGTTGCT 4079  
3961 TCCAAACAAATGAGGATCAGTCTGAAACCCAGGGAGTTGCTGTAGTACAGAGAAATG 4020  
|||||  
4080 TCCAAACAAATGAGGATCAGTCTGAAACCCAGGGAGTTGCTGTAGTACAGAGAAATG 4139  
4021 GTTTCAGATGATGAGAAAGAGAGAGGCGCTTGGAGAAATATATCAAGAGAGCAAGC 4080  
|||||  
4140 GTTTCAGATGATGAGAAAGAGAGAGGCGCTTGGAGAAATATATCAAGAGAGCAAGC 4199  
4081 ATGATTTCAACTTATGATGAGAGAGAGAGGCTTGGAGAGTGAAGAAACAGGCTCTGAA 4140  
|||||  
4200 ATGATTTCAACTTATGATGAGAGAGAGAGGCTTGGAGAGTGAAGAAACAGGCTCTGAA 4259  
4141 GACTGCTCAGGGCTATCTCTCAGAGTGCATTTTAACCACTCAGCAGAGGATACCATG 4200  
|||||  
4260 GACTGCTCAGGGCTATCTCTCAGAGTGCATTTTAACCACTCAGCAGAGGATACCATG 4319  
4201 CAACATTAACCTGATTAAGCTCCAGCAGGAAATGCGTGAATGAAAGTGTGTAGAACAG 4260  
|||||  
4320 CAACATTAACCTGATTAAGCTCCAGCAGGAAATGCGTGAATGAAAGTGTGTAGAACAG 4379  
4261 CATGGAGGACGAGCTCTTAACAGCTACCTCCATCATATAGTACGCTTCTGCCCCCTGAG 4320  
|||||  
4380 CATGGAGGACGAGCTCTTAACAGCTACCTCCATCATATAGTACGCTTCTGCCCCCTGAG 4439  
4321 GACCTGCAAAATCCAGAACAAACACATCAGAAAAAGCAGTAACTTACCTCAGAAAAAGT 4380  
|||||  
4440 GACCTGCAAAATCCAGAACAAACACATCAGAAAAAGCAGTAACTTACCTCAGAAAAAGT 4499  
4381 AGTGAATACCTTATTAAGCCAGAAATCCAGAGGCTTTCTGCTGACAAATTTGAGGTGCT 4440  
|||||  
4500 AGTGAATACCTTATTAAGCCAGAAATCCAGAGGCTTTCTGCTGACAAATTTGAGGTGCT 4559  
4441 GCAGATAGTTCTAACAGTAAATAAAGAAAGCAGAGTGAAGGCTCATCCCTCTTAAA 4500  
|||||  
4560 GCAGATAGTTCTAACAGTAAATAAAGAAAGCAGAGTGAAGGCTCATCCCTCTTAAA 4619  
4501 TGCCCATCATTAAGATATAGTGTATACATGACAGATTGCTCTGGAGTCTTCAAGATAGA 4560  
|||||  
4620 TGCCCATCATTAAGATATAGTGTGTATACATGACAGATTGCTCTGGAGTCTTCAAGATAGA 4679  
4561 AACTACCATCTCAGAGAGAGTCAATTAAGTGTGTATGTGTGAGAGACACACACTGGAA 4620  
|||||  
4680 AACTACCATCTCAGAGAGAGTCAATTAAGTGTGTATGTGTGAGAGACACACACTGGAA 4739  
4621 GAGTCTGGGCGACAGATTTGCGGAAACATTTACTTCCCAAGGCAAGATCTGAGAGGA 4680  
|||||  
4740 GAGTCTGGGCGACAGATTTGCGGAAACATTTACTTCCCAAGGCAAGATCTGAGAGGA 4799  
4681 ACCCCTTACCTGGAATCTGGAATCAGCCTTCTCTCTGATGAGACCTGATGATCTTCT 4740  
|||||  
4800 ACCCCTTACCTGGAATCTGGAATCAGCCTTCTCTCTGATGAGACCTGATGATCTTCT 4859  
4741 GAAGACAGAGCCCAAGAGTCACTGCTGTGGCAACATACATCTTCAACCTCTGCAATG 4800  
|||||

Db	4860	GAAGACAGAGGCCCCAGAGTGCATCGTGTGGTGGCAACATACCATCTTCAACCTCGCATGTG	49119
QY	4801	AAAGTTCCTCCCAATTGAAAGTTCGACGAATCTGCCCCAGAGTCCAGCTGCTGCTCACTACT	48606
Db	4920	AAAGTTCCTCCCAATTGAAAGTTCGACGAATCTGCCCCAGAGTCCAGCTGCTGCTCACTACT	49797
QY	4861	GATCATGCTGGGTATTAATGCAATGGAGAAAGTGTGAGCAGAGGAGAAAGCCAGAAATTGACA	49200
Db	4980	GATCATGCTGGGTATTAATGCAATGGAGAAAGTGTGAGCAGAGGAGAAAGCCAGAAATTGACA	50399
QY	4921	GCTTCAACAGAGAAAGGCTCAACAAAAGAAATGTCATGTGGTGGTGGCTCAACCCACAGA	49808
Db	5040	GCTTCAACAGAGAAAGGCTCAACAAAAGAAATGTCATGTGGTGGTGGCTCAACCCACAGA	50999
QY	4981	GAATTTATGCTCGTGTACAAAGTTGGCCAGAAACACCAATCACTTTAACTAAATCTAAT	50400
Db	5100	GAATTTATGCTCGTGTACAAAGTTGGCCAGAAACACCAATCACTTTAACTAAATCTAAT	51599
QY	5041	ACTGAGAGAGTACTCATGTTGTTATGAAAAACAGTGTGAGTTGTGTGTAACGGACA	51000
Db	5160	ACTGAGAGAGTACTCATGTTGTTATGAAAAACAGTGTGAGTTGTGTGTAACGGACA	52199
QY	5101	CTGAATATTTTTCATGAGAAATGGCGGGAGAAATAGGGAGTAACTACTATTTCTGGGTGAC	51606
Db	5220	CTGAATATTTTTCATGAGAAATGGCGGGAGAAATAGGGAGTAACTACTATTTCTGGGTGAC	52799
QY	5161	CAGTCTATTTAAAGAAAGAAAAATCTGCTGATGAGATGATTTTGAAGTCAGACGAGATGTG	52200
Db	5280	CAGTCTATTTAAAGAAAGAAAAATCTGCTGATGAGATGATTTTGAAGTCAGACGAGATGTG	53399
QY	5221	GTCATATGAGAAACCACCAAGGTCCAAAGCGAGCAAGAGATATCCAGAGACAGAAAGATC	52800
Db	5340	GTCATATGAGAAACCACCAAGGTCCAAAGCGAGCAAGAGATATCCAGAGACAGAAAGATC	53999
QY	5281	TTCAAGGGGGCTAGAAATCTGTGTGATAGGGCCCTTCCACCAATGAGCCACAGATCAACTG	53400
Db	5400	TTCAAGGGGGCTAGAAATCTGTGTGATAGGGCCCTTCCACCAATGAGCCACAGATCAACTG	54599
QY	5341	GAATGAGTGTATACAGCTGTGTGTCTCTGTGTGTGTAAGAGAGCTTTGATCATCTACCCCT	54000
Db	5460	GAATGAGTGTATACAGCTGTGTGTCTCTGTGTGTGTAAGAGAGCTTTGATCATCTACCCCT	55199
QY	5401	GGCACAGGTGTCCACCCCAATTTGTGTTGTGTCAGCCAGATGCTTGGACAGAGCAATGGC	54606
Db	5520	GGCACAGGTGTCCACCCCAATTTGTGTTGTGTCAGCCAGATGCTTGGACAGAGCAATGGC	55799
QY	5461	TTTCATGCAATTTGGGACAGATGTGTGAGGCACCTGTGTGTGAGCCCAAGATGGGTGTGAC	55200
Db	5580	TTTCATGCAATTTGGGACAGATGTGTGAGGCACCTGTGTGTGAGCCCAAGATGGGTGTGAC	56399
QY	5521	AGTTAGCACTCTACAGATGTCAGAGAGTGGACACTACTAGATATCCCAATGATCCCAAC	55800
Db	5640	AGTTAGCACTCTACAGATGTCAGAGAGTGGACACTACTAGATATCCCAATGATCCCAAC	56999
QY	5581	AGCCACTATC 5589	
Db	5700	AGCCACTATC 5708	
RESULT 3			
LOCUS	AR033056	5711 bp	DNA
DEFINITION	Sequence 1 from patent US 5869245.	linear	PAT 29-SEP-1999
ACCESSION	AR033056		
VERSION	AR033056.1	GI:5948661	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5711)		
AUTHORS	Yeung,A.T.		
TITLE	Mismatch endonuclease and its use in identifying mutations in targeted polynucleotide strands		

JOURNAL FEATURES		Patent: US 5869245-A 1 09-FEB-1999:		Location/Qualifiers	
source		1. .5711		/organism="unknown"	
BASE COUNT	1956 a	1099 c	1274 g	1382 t	
ORIGIN					
Query Match	99.9%;	Score 5585.8;	DB 6;	Length 5711;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 5587;	Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0
QY	1	ATGATTTATCTGCTCTTCCGCTTAAAGACGACCAAAATGTCTATTATGCTATGCGA	60		
DB	120	ATGATTTATCTGCTCTTCCGCTTAAAGACGACCAAAATGTCTATTATGCTATGCGA	179		
QY	61	ATCTTAGAGTGCCTCATCTGCTGAGTTGATCAAGGAACTGCTGCCCAAGTGTGC	120		
DB	180	ATCTTAGAGTGCCTCATCTGCTGAGTTGATCAAGGAACTGCTGCCCAAGTGTGC	239		
QY	121	CACATATTTTGCAAATTTTGCATGCTGAACCTTCTCAACACAGAAAGGGCTTCACAG	180		
DB	240	CACATATTTTGCAAATTTTGCATGCTGAACCTTCTCAACACAGAAAGGGCTTCACAG	299		
QY	181	TGTCCTTTATGTAGAAATGATTAACCAAAAGAGCTTCAAGAAAGTACGATTTAGT	240		
DB	300	TGTCCTTTATGTAGAAATGATTAACCAAAAGAGCTTCAAGAAAGTACGATTTAGT	359		
QY	241	CACCTGTTGGAAGACCTATGTAATTCATTTGCTTTGACCTTGACACAGGTTTGAG	300		
DB	360	CACCTGTTGGAAGACCTATGTAATTCATTTGCTTTGACCTTGACACAGGTTTGAG	419		
QY	301	TATGCAAAACAGCTATATTTTGCAGAAAAGGAAATTAACCTCTCTGAACATCTAAAGAT	360		
DB	420	TATGCAAAACAGCTATATTTTGCAGAAAAGGAAATTAACCTCTCTGAACATCTAAAGAT	479		
QY	361	GAACTTTTATCATCCAAAGTATGGGCTACAGAAACCGTCCCAAGACTTTCACAGAT	420		
DB	480	GAACTTTTATCATCCAAAGTATGGGCTACAGAAACCGTCCCAAGACTTTCACAGAT	539		
QY	421	GAACCCGAAATTCCTTCCCTTGAGGAAACCGTCCAGTGTCAACCTCTAACCTGGA	480		
DB	540	GAACCCGAAATTCCTTCCCTTGAGGAAACCGTCCAGTGTCAACCTCTAACCTGGA	599		
QY	481	ACTGTGAGACCTGTGAGGACAAAGGAGGATACCACTCAAAAAGCGTCTTACATTT	540		
DB	600	ACTGTGAGACCTGTGAGGACAAAGGAGGATACCACTCAAAAAGCGTCTTACATTT	659		
QY	541	GAATTTGGATCTGATTTCTTCTGAAAGATACCGTTAATTAAGCAACTTATTGACGTGGGA	600		
DB	660	GAATTTGGATCTGATTTCTTCTGAAAGATACCGTTAATTAAGCAACTTATTGACGTGGGA	719		
QY	601	GATCAAGAATTTCTTCAAAATCAACCCCTCAAGAACACAGGATGAATCAGTTTGCATCT	660		
DB	720	GATCAAGAATTTCTTCAAAATCAACCCCTCAAGAACACAGGATGAATCAGTTTGCATCT	779		
QY	661	GCAAAAAAGCGTCTTGTGAATTTCTGAGACGAGTGTAAACAAATTAAGAACATCATCA	720		
DB	780	GCAAAAAAGCGTCTTGTGAATTTCTGAGACGAGTGTAAACAAATTAAGAACATCATCA	839		
QY	721	CCCAAGTAATATGATTTTGAACACCACTGAGAAAGCGTGCAGTGAGAGGCATCCAGAAAG	780		
DB	840	CCCAAGTAATATGATTTTGAACACCACTGAGAAAGCGTGCAGTGAGAGGCATCCAGAAAG	899		
QY	781	TATCAGGTAATTTCTGTTTCAAACTGTGATGTGAGAGCCATGTGGACAAATTAATCTATGCC	840		
DB	900	TATCAGGTAATTTCTGTTTCAAACTGTGATGTGAGAGCCATGTGGACAAATTAATCTATGCC	959		
QY	841	AGCTCATTTACAGCATGAGAACAGCGATTTATTACTCAATAAGACAGATGATTTAGAA	900		
DB	960	AGCTCATTTACAGCATGAGAACAGCGATTTATTACTCAATAAGACAGATGATTTAGAA	1019		
QY	901	AAGCTGAATTTCTGTAATTAAGCAACAGCGCTGAGTACGAAAGGACCAATTAACGA	960		



Db 1020 AAGGCTGAATCTGTAATAAAAGCAAAACGCTGGCTTAGCAAGAGCAACATATAACAGA 1079  
QY 961 TGGGCTGGAAGTAAAGAAACATGTAAATGATGGCGGACCTCCAGCAGCAAGAAAAAGATA 1020  
Db 1080 TGGGCTGGAGTAAGAAACATGTAAATGATGGCGGACCTCCAGCAGCAAGAAAAAGATA 1139  
QY 1021 GATCGAATGCTGATCCCTGTGTGAGAGAAAAAAGTAATGGAATAGCAGAACTGCCATGC 1080  
Db 1140 GATCGAATGCTGATCCCTGTGTGAGAGAAAAAAGTAATGGAATAGCAGAACTGCCATGC 1199  
QY 1081 TCAGAGAACTCTAGAGATCTAGAGATCTGATCTCTGATTAACACTAAATAGCAGATTGAG 1140  
Db 1200 TCAGAGAACTCTAGAGATCTAGAGATCTGATCTCTGATTAACACTAAATAGCAGATTGAG 1259  
QY 1141 AAAGTAATGAGTGGTTTCCAGAGTGAATGAACTGTAGGTCTGATGCTCAGATGAT 1200  
Db 1280 AAAGTAATGAGTGGTTTCCAGAGTGAATGAACTGTAGGTCTGATGCTCAGATGAT 1319  
QY 1201 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGTAATGGACGTTCTAAATAGTAGAT 1260  
Db 1320 GGGGAGTCTGAATCAAAATGCCAAAGTAGCTGATGTAATGGACGTTCTAAATAGTAGAT 1379  
QY 1281 GAATATCTGCTTCTTCAGAGAAATAGACTTACTGGCCAGTATCTCTATGAGCTTTA 1320  
Db 1380 GAATATCTGCTTCTTCAGAGAAATAGACTTACTGGCCAGTATCTCTATGAGCTTTA 1439  
QY 1321 ATATGTAAAGTGAAGAGTTCCTACTCCAAATCAGTAGAGATTAATTTAGAGCAAAATA 1380  
Db 1440 ATATGTAAAGTGAAGAGTTCCTACTCTCCAAATCAGTAGAGATTAATTTAGAGCAAAATA 1499  
QY 1381 TTTGGAAAAACCTATCGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAATCTGAAT 1440  
Db 1500 TTTGGAAAAACCTATCGAAGAGGCAAGCCCTCCCACTTAAGCCATGTAATCTGAAT 1559  
QY 1441 CTAATTAAGAGCATTTTGTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAT 1500  
Db 1560 CTAATTAAGAGCATTTTGTCTAGAGCCACAGATTAATCAAGAGGCTCCCTCACAAT 1619  
QY 1501 AAATTAAGAGCATTAAGAGAGACCTACATGAGCCTTCACTGAGATTTTATCAAGAA 1560  
Db 1620 AAATTAAGAGCATTAAGAGAGACCTACATGAGCCTTCACTGAGATTTTATCAAGAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAGACCTCCTGAATGATTAATCAGGGAATTAACCAAGCGAG 1620  
Db 1680 GCAGATTTGGCAGTTCAAAAGACCTCCTGAATGATTAATCAGGGAATTAACCAAGCGAG 1739  
QY 1621 CAGATGCTCAAGTGAATGAATTTACTATAGTGGTCATGAGATTAACCAAGAGTGAT 1680  
Db 1740 CAGATGCTCAAGTGAATGAATTTACTATAGTGGTCATGAGATTAACCAAGAGTGAT 1799  
QY 1681 TCTATTCAGAAATGAGAAAAATCTTAACCAATAGAACTCTGAAAAAAGATCTGCTTC 1740  
Db 1800 TCTATTCAGAAATGAGAAAAATCTTAACCAATAGAACTCTGAAAAAAGATCTGCTTC 1859  
QY 1741 AAAACGAAGCTGAACCTATTAAGCAGCAGTATTAAGCAATATGGAATCTGAATTAATATC 1800  
Db 1860 AAAACGAAGCTGAACCTATTAAGCAGCAGTATTAAGCAATATGGAATCTGAATTAATATC 1919  
QY 1801 CACATTTAAAAAGCACTAAAAAGATAGCTGAGAGAGAACTCTTCTACCGAGCATAT 1860  
Db 1920 CACATTTAAAAAGCACTAAAAAGATAGCTGAGAGAGAACTCTTCTACCGAGCATAT 1979  
QY 1861 CATGGCTTGAACCTAGTAGTGAATTAAGCCACCTAATTTGTAAGTGAATGCA 1920  
Db 1980 CATGGCTTGAACCTAGTAGTGAATTAAGCCACCTAATTTGTAAGTGAATGCA 2039  
QY 1921 ATTGATGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCGCAGTC 1980  
Db 2040 ATTGATGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCGCAGTC 2099  
QY 1981 AGGCACACAGAAACCTACAACTCATGGAAGGTAAAGAACTGCAACCTGAGCCAAAGAG 2040  
Db 2100 AGGCACACAGAAACCTACAACTCATGGAAGGTAAAGAACTGCAACCTGAGCCAAAGAG 2159

QY 2041 AGTAACACCAATGAAACAGACAGATAAAAGACATGACAGTACTTTCCAGAGCT 2100  
Db 2160 AGTAACACCAATGAAACAGACAGATAAAAGACATGACAGTACTTTCCAGAGCT 2219  
QY 2101 AAGTTAAACAAATGACACCTGGTTCTTTACTAGTGTCAAAATACCAAGTGAACCTTAAGAA 2160  
Db 2220 AAGTTAAACAAATGACACCTGGTTCTTTACTAGTGTCAAAATACCAAGTGAACCTTAAGAA 2279  
QY 2161 TTTGTCAATCTAGCTTCCAGAGAGAAAAAGAAAGAACTTGAACAGTTAAAGTG 2220  
Db 2280 TTTGTCAATCTAGCTTCCAGAGAGAAAAAGAAAGAACTTGAACAGTTAAAGTG 2339  
QY 2221 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGCTTTGCAACT 2280  
Db 2340 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGCTTTGCAACT 2339  
QY 2281 GAAAGATCTGTAGAGAGTAGAGATTTCAATTTGTAAGTGTATGATGAGCACTGAG 2340  
Db 2400 GAAAGATCTGTAGAGAGTAGAGATTTCAATTTGTAAGTGTATGATGAGCACTGAG 2459  
QY 2341 GAAAGATCTGTAGAGAGTAGAGATTTCAATTTGTAAGTGTATGATGAGCACTGAG 2400  
Db 2460 GAAAGATCTGTAGAGAGTAGAGATTTCAATTTGTAAGTGTATGATGAGCACTGAG 2519  
QY 2401 TGTGTAGTCAAGTGTGACAGCATTTGAAAAACCCCAAGGAGCTAATTCATGTTGCCAAA 2460  
Db 2520 TGTGTAGTCAAGTGTGACAGCATTTGAAAAACCCCAAGGAGCTAATTCATGTTGCCAAA 2579  
QY 2461 GATTAATAGAAATGACACAGAGGCTTTAAGTATCCATTTGGAGATGAAGTTAACACAGT 2520  
Db 2580 GATTAATAGAAATGACACAGAGGCTTTAAGTATCCATTTGGAGATGAAGTTAACACAGT 2639  
QY 2521 CGGGAACAGCATTAAGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGCCAATTA 2580  
Db 2640 CGGGAACAGCATTAAGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGCCAATTA 2639  
QY 2581 TTTCAAGGTTTCAAAAGCGCAGTCAATTTGCTGTTTCAATTCAGAAATGCGAAGAG 2640  
Db 2700 TTTCAAGGTTTCAAAAGCGCAGTCAATTTGCTGTTTCAATTCAGAAATGCGAAGAG 2759  
QY 2641 GAATGTCACATTTCTGCCCCACTGTGGGCTTTAAAGAAACAAAGTCAAGTCACT 2700  
Db 2760 GAATGTCACATTTCTGCCCCACTGTGGGCTTTAAAGAAACAAAGTCAAGTCACT 2819  
QY 2701 TTTGAATGGAACAAAGAGAAAGAAATCAAGAAAGAAATGATGATTAATCAACCTGTA 2760  
Db 2820 TTTGAATGGAACAAAGAGAAAGAAATCAAGAAAGAAATGATGATTAATCAACCTGTA 2879  
QY 2761 CAGACAGTTAATATCACTGCAAGGCTTCTGTTGTTGTCAGAAAGATTAAGCCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCACTGCAAGGCTTCTGTTGTTGTCAGAAAGATTAAGCCAGTTGAT 2939  
QY 2821 AATGCCAAATGATATCAAGAGAGGCTGATGTTTGTATCATCTCAGTCAAGAGC 2880  
Db 2940 AATGCCAAATGATATCAAGAGAGGCTGATGTTTGTATCATCTCAGTCAAGAGC 2939  
QY 2881 AAGCAATGCAAGCTATTAATCAATTAAGTGAAGCTTTTACAAACCCATATCAGTATA 2940  
Db 3000 AAGCAATGCAAGCTATTAATCAATTAAGTGAAGCTTTTACAAACCCATATCAGTATA 3059  
QY 2941 CCACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATTAAGAAAAATCTGCTAGAG 3000  
Db 3060 CCACCACTTTTCCCATCAAGTCAATTTGTTAAACTAATTAAGAAAAATCTGCTAGAG 3119  
QY 3001 GAAAACTTTGAGGAACATTCATGACCTGAAAGAGAAATGGAACATTCCTCA 3060  
Db 3120 GAAAACTTTGAGGAACATTCATGACCTGAAAGAGAAATGGAACATTCCTCA 3179  
QY 3061 AGTACAGTGAACAAATTAAGCCGTAATTAAGTGAAGAAATGTTTAAAGAAAGCCAGC 3120  
Db 3180 AGTACAGTGAACAAATTAAGCCGTAATTAAGTGAAGAAATGTTTAAAGAAAGCCAGC 3239



QY 3121 TCACCAATATTAAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3180  
| | | | |  
Db 3240 TCACCAATATTAAAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3299  
| | | | |  
QY 3181 ATAGGTTCAGTGAAGAAACATTCAGCAGAACTAGTGAAGAAAGAGGGCCAAATTTG 3240  
| | | | |  
Db 3300 ATAGGTTCAGTGAAGAAACATTCAGCAGAACTAGTGAAGAAAGAGGGCCAAATTTG 3359  
| | | | |  
QY 3241 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3300  
| | | | |  
Db 3360 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3419  
| | | | |  
QY 3301 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3360  
| | | | |  
Db 3420 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3479  
| | | | |  
QY 3361 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3420  
| | | | |  
Db 3480 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3539  
| | | | |  
QY 3421 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3480  
| | | | |  
Db 3540 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3599  
| | | | |  
QY 3481 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3540  
| | | | |  
Db 3600 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3659  
| | | | |  
QY 3541 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3600  
| | | | |  
Db 3660 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3719  
| | | | |  
QY 3601 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3660  
| | | | |  
Db 3720 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3779  
| | | | |  
QY 3661 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3720  
| | | | |  
Db 3780 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3839  
| | | | |  
QY 3721 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3780  
| | | | |  
Db 3840 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3899  
| | | | |  
QY 3781 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3840  
| | | | |  
Db 3900 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3959  
| | | | |  
QY 3841 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3900  
| | | | |  
Db 3960 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4019  
| | | | |  
QY 3901 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 3960  
| | | | |  
Db 4020 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4079  
| | | | |  
QY 3961 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4020  
| | | | |  
Db 4080 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4139  
| | | | |  
QY 4021 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4080  
| | | | |  
Db 4140 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4199  
| | | | |  
QY 4081 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4140  
| | | | |  
Db 4200 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4259  
| | | | |  
QY 4141 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4200  
| | | | |  
Db 4260 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4319  
| | | | |  
QY 4201 ATAGGTTCAGTGAAGAAAGGTTTGCACCTGAGGCTATTAACAAAGTCTCTGGA 4260  
| | | | |

Db 4320 CAACATTAACCTGATAAGAGTCCAGCAGAAATGGCTGAACTAGAACTGTGTAACAG 4379  
| | | | |  
QY 4261 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4320  
| | | | |  
Db 4380 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4439  
| | | | |  
QY 4321 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4380  
| | | | |  
Db 4440 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4499  
| | | | |  
QY 4381 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4440  
| | | | |  
Db 4500 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4559  
| | | | |  
QY 4441 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4500  
| | | | |  
Db 4560 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4619  
| | | | |  
QY 4501 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4560  
| | | | |  
Db 4620 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4679  
| | | | |  
QY 4561 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4620  
| | | | |  
Db 4680 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4739  
| | | | |  
QY 4621 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4680  
| | | | |  
Db 4740 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4799  
| | | | |  
QY 4681 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4740  
| | | | |  
Db 4800 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4859  
| | | | |  
QY 4741 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4800  
| | | | |  
Db 4860 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4919  
| | | | |  
QY 4801 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4860  
| | | | |  
Db 4920 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4979  
| | | | |  
QY 4861 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4920  
| | | | |  
Db 4980 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5039  
| | | | |  
QY 4921 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 4980  
| | | | |  
Db 5040 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5099  
| | | | |  
QY 4981 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5040  
| | | | |  
Db 5100 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5159  
| | | | |  
QY 5041 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5100  
| | | | |  
Db 5160 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5219  
| | | | |  
QY 5101 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5160  
| | | | |  
Db 5220 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5279  
| | | | |  
QY 5161 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5220  
| | | | |  
Db 5280 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5339  
| | | | |  
QY 5221 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5280  
| | | | |  
Db 5340 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5399  
| | | | |  
QY 5281 CATGGAGCAGCCTCTTACAGCTACATACCTCTCATATTAAGTACTCTCTGCTGAG 5340  
| | | | |

```

Db      5400 TTCAGGGGGCTAGAAATCTGTTGCTATGAGGGCCCTTCCACCAAGATGCCACAGATCAACTG 5459
Oy      5341 GAATGATGATGACAGCTGTGCTGCTCTTCTGTGTGTAAGAGAGCTTTTCATTCATCCACCTT 5400
        |||||||
Db      5460 GAATGATGATGACAGCTGTGCTGCTCTTCTGTGTGTAAGAGAGCTTTTCATTCATCCACCTT 5519
Oy      5401 GGCACAGGTGTCACCCCAATTTGTTGTTGTGTCAGACCAAGATGCTGACAGAGACAATGCG 5460
        |||||||
Db      5520 GGCACAGGTGTCACCCCAATTTGTTGTTGTGTCAGACCAAGATGCTGACAGAGACAATGCG 5579
Oy      5461 TTCATGCAATTTGGGCGAGATGTGTGAGAGCACTGCTGCTGACAGAGAGTGGTGTGAC 5520
        |||||||
Db      5560 TTCATGCAATTTGGGCGAGATGTGTGAGAGCACTGCTGCTGACAGAGAGTGGTGTGAC 5639
Oy      5521 AGTGTAGCACTCTACAGTGCAGAGAGCTGACACCTACCTGATACCCAGATCCGCCAC 5580
        |||||||
Db      5640 AGTGTAGCACTCTACAGTGCAGAGAGCTGACACCTACCTGATACCCAGATCCGCCAC 5699
Oy      5561 AGCCACTAC 5589
        |||||||
Db      5700 AGCCACTAC 5708

RESULT 4
HS014680 5711 bp mRNA linear PRI 10-JUN-2002
LOCUS Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA,
DEFINITION complete cds.
ACCESSION U14680
VERSION U14680.1 GI:555931
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5711)
AUTHORS Miki,Y., Swensen,J., Shattuck-Eidens,D., Futreal,P.A., Harshman,K.,
            Tavtigian,S., Liu,Q., Cochran,C., Bennett,L.M., Ding,W., Bell,R.,
            Rosenthal,J., Hussey,C., Tran,T., McClure,M., Frye,C., Hattler,T.,
            Phelps,R., Haugen-Strano,A., Katcher,H., Yakumo,K., Gholami,Z.,
            Shaffer,D., Stone,S., Bayer,S., Wray,C., Bogden,R., Dayananth,P.,
            Ward,J., Tonin,P., Narod,S., Bristow,P.K., Norris,F.H.,
            Helvering,L., Morrison,P., Rostock,P., Lai,M., Barret,J.C.,
            Lewis,C., Neuhausen,S., Cannon-Albright,L., Goldgar,D., Wiseman,R.,
            Kamb,A. and Skolnick,M.H.
            A strong candidate for the breast and ovarian cancer susceptibility
            gene BRCA1
            Science 266 (5182), 66-71 (1994)
TITLE gene BRCA1
JOURNAL Science 266 (5182), 66-71 (1994)
MEDLINE 95025896
PUBMED 7545954
REFERENCE 2 (bases 1 to 5711)
AUTHORS Skolnick,M.H.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1994) Mark H. Skolnick, Myriad Genetics Inc. and
            the University of Utah, 421 Wakara Way, Suite 201, Salt Lake City,
            UT 84108, USA
FEATURES
source Location/Qualifiers
    1..5711
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="17"
        /map="17q21: spans D17S855"
        /note="For sequence of alternatively spliced exon 4, see
            Genbank Accession Number U15595"
    1..5711
        /gene="BRCA1"
    1..119
        /gene="BRCA1"
    1..100
        /gene="BRCA1"
    /number=1
    101..199
        /gene="BRCA1"
    exon
    120..253
        /gene="BRCA1"
    /number=3
    254..331
        /gene="BRCA1"
    /number=5
    332..420
        /gene="BRCA1"
    /number=6
    421..560
        /gene="BRCA1"
    /number=7
    561..665
        /gene="BRCA1"
    /number=8
    666..712
        /gene="BRCA1"
    /number=9
    713..788
        /gene="BRCA1"
    /number=10
    789..4215
        /gene="BRCA1"
    /number=11
    4216..4302
        /gene="BRCA1"
    /number=12
    4303..4476
        /gene="BRCA1"
    /number=13
    4477..4603
        /gene="BRCA1"
    exon
    /number=2
    120..5711
        /gene="BRCA1"
        /note="influences susceptibility to breast and ovarian
            cancer"
        /codon_start=1
        /product="breast and ovarian cancer susceptibility"
        /protein_id="AA73985.1"
        /db_xref="GI:555932"
        /translation="WDLSALRVEVQVAVNAMOKILDEPCLEILKEPVSTKCDHIFC
            KFCMLKLNQKGRSQCPKCNKDTKRSLOSTFESQVLELKLITAFDLGLEA
            NSYNFAKKEENSPKHLKDEVSITQSKYRNRAKRLLOSPEENPSLETSTLSVLNIG
            TVRLTQKRIQPKTSVYIELGSDSEEDYRNKATYGVSDQELQITPGTDEISL
            DSAKAACSEFSETVTNTEHHQPSNNLNTETKRAARHPEKYGSSVSLAHPECGT
            NTHASLSQHENSILLTRDRNVKAEKCKSKOPGLARSHNRMRGSKTCDRNP
            STEKVDIADNPCEKEMNKQKIPCSNPDPDPDVTWITLNSIQVNWMPRSBEL
            LGSDDSHDGESENKAVADVLDVLEVDERTSGSEKIDLASDPHEALICKSRVSK
            SVESNIEDKIFGKTYRKKASLPNLSHVTENLIGAFVTEPQIIQERPLNKLKRRRP
            TSGLPEDPIKRADLAVOKTPMINOGTQNGQVYMNITNSGHEKTKGDSIQNEK
            NPNPLESEKSAKTAEPISISSNMELELNHNSAKRNRLRRKSRSTRIHALE
            LVSRNLSPNCTELOIDSCSSSEIKKKKYNOMPVHRSNLOIMECKEPATGAKSN
            KPNPOTSRHSDPEPELKLINAGSPTKCSNTSELKEPNPSLPREKKELETYK
            SNMADPKDMLSGERYLQTERSVSSISLVPETDGTQESTLSLEVSTLGAETEP
            NKCVSCAFAENPKGLHGCSDKRNNDTEGKYPGLGHEVHNSRETSLMESELDQY
            LQNTFKYSKRSQAFSPNPGNAEECAFFSAHSGSLKQSPKYTFEEOKEENQKNE
            SNIKPEVQVNTITAGEPVYQKORQVDNAKCSIKGSRFCSSQFQGNETCLTPNKG
            LLQNPYRIPLPEPKSPFKCKKNLLEENPEEHSPEREMENPSTVSTISRNN
            IRENVKASSSNINENQSSNPVGSINETSDEMDIQALELRNRPKNAMLRIGV
            LQPEYTKOSLPGSMCKHPEIKQETEEVQVNTDPSPTLSDNLEQPMGSSSAQVC
            SETPDLLDDELKEDTSPFAENDIKESAVFSKSVQGLSRSPSPPTHLAAGYR
            GAKLSESEENLSEDELPQFHLFGKVAKNIPSOSTRSHVATGLSTNDENLIS
            LKNSLNDSCNOVILAKASQEHLSSEETKCSALPSQSELEDLTANTNODPFLGS
            SKOMRHSQSGVGLSDKELYSODEERGTLLENNOEQSDNLSGAAAGCSETSV
            SEDQSGSSQSDILTTQORDMQNLIKIQEAMLEAVAEQHSQSPNSPRTIDS
            SAELELRNPDESTSEKAVLUSQKSEFPISQNPGLSADPFVSABDSITKKNPEVE
            KSSPSKCPSLDENYMHSCGSLQNNRNPQOEILKVDVEEQLSESGHDLTESY
            LPRLDEGTPYLEGISLFSDDPSDSEDAEPASAVANGVIPSSTSLAKVPQKLVSY
            AQSPAAATTTAGYNAMEESVSRKEPDLASTERYVRKMSAVVSGITPEEFLYKF
            ARKHHTITNTLITEETHVYKRTAEFCERTLKYFLKIGGKMWVSYFVTDISIKR
            KMLNHDPEVRGVDVNNRNGHOPRARESODRKFRGLTICGCPFTNPTDQLEMY
            OLCASVYKELSSPTLCTGVPIYVGPDMATENDNGHAIQMCAPVYTRVAVLDSV
            ALYQCELDITLIPITPSHY"

```

```

/ gene="BRCA1"
/ number=14
4604 . 4794
/ gene="BRCA1"
/ number=15
4795 . 5105
/ gene="BRCA1"
/ number=16
5106 . 5193
/ gene="BRCA1"
/ number=17
5194 . 5273
/ gene="BRCA1"
/ number=18
5274 . 5310
/ gene="BRCA1"
/ number=19
5311 . 5396
/ gene="BRCA1"
/ number=20
5397 . 5451
/ gene="BRCA1"
/ number=21
5452 . 5526
/ gene="BRCA1"
/ number=22
5527 . 5586
/ gene="BRCA1"
/ number=23
5587 . 5711
/ gene="BRCA1"
/ number=24

BASE COUNT      1956 a      1099 c      1274 g      1382 t
ORIGIN
Query Match      99.98; Score 5585.8; DB 9; Length 5711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  ATGGATTATCTGCTCTGCTGCTGGAAGATCAAAATGTCATTATGCTATGCGAAA 60
DB      120 ATGGATTATCTGCTCTGCTGCTGGAAGATCAAAATGTCATTATGCTATGCGAAA 179
QY      61 ATCTTAGAGTGTCCATCTGTCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGAC 120
DB      180 ATCTTAGAGTGTCCATCTGTCTGAGTTGATCAAGAACCTGTCTCCAAAGTGTGAC 239
QY      121 CACATATTTTGCATAATTTTGCATGCTGAACCTTCTCAACGAGAAAGGCTTCACAG 180
DB      240 CACATATTTTGCATAATTTTGCATGCTGAACCTTCTCAACGAGAAAGGCTTCACAG 239
QY      181 TGTCTTTATGTAGATGATATACCAAAAGAGCTTACAAAGATAGCATTTAGT 240
DB      300 TGTCTTTATGTAGATGATATACCAAAAGAGCTTACAAAGATAGCATTTAGT 339
QY      241 CAATCTGTGAGAGCTATTAAGAAATCATTTGTCTTTCAGCTTGACACAGGTTTGAG 300
DB      360 CAATCTGTGAGAGCTATTAAGAAATCATTTGTCTTTCAGCTTGACACAGGTTTGAG 419
QY      301 TATGCAAAACAGCTATTAATTTTGCAGAAAGGAAATTAATCTCTCCTGAACATCTAAAGAT 360
DB      420 TATGCAAAACAGCTATTAATTTTGCAGAAAGGAAATTAATCTCTCCTGAACATCTAAAGAT 479
QY      361 GAAGTTTATCATCAAAAGATATGGCTACAGAAACGCTGCCAAAGAGCTTACAGAGT 420
DB      480 GAAGTTTATCATCAAAAGATATGGCTACAGAAACGCTGCCAAAGAGCTTACAGAGT 539
QY      421 GAACCCGAAATCTCTCTTGAGAGAACAGCTCAGTGTCCAACTCTTAACCTTGA 480
DB      540 GAACCCGAAATCTCTCTTGAGAGAACAGCTCAGTGTCCAACTCTTAACCTTGA 539
QY      481 ACTGTGAGAACTGTGAGAGAACAGCGGATACAACTCAAAAGAGCTGTCTACAT 540

```

```

DB      600 ACTGTGAGAACTGTGAGAGAACAGCGGATACAACTCAAAAGAGCTGTCTACAT 659
QY      541 GAATGGAGATCTGATTTCTTGAGAGATACCGTTAATAAGCAACTATTCAGAGTGGGA 600
DB      660 GAATGGAGATCTGATTTCTTGAGAGATACCGTTAATAAGCAACTATTCAGAGTGGGA 719
QY      601 GATCAGAAATTTGATCAAAATCACCCCTCAAGAACAGGAGTGAATCAGTTGGATTC 660
DB      720 GATCAGAAATTTGATCAAAATCACCCCTCAAGAACAGGAGTGAATCAGTTGGATTC 779
QY      661 GCAAAAAGGCTCTGTGATTTTCTGAGACGAGATGTACAAATATCTGAACATCA 720
DB      780 GCAAAAAGGCTCTGTGATTTTCTGAGACGAGATGTACAAATATCTGAACATCA 839
QY      721 CCCAGTAAATGATTTGAACACACAGAGAGCGTCACTGAGAGGATCCAGAAAG 780
DB      840 CCCAGTAAATGATTTGAACACACAGAGAGCGTCACTGAGAGGATCCAGAAAG 899
QY      781 TATCAGGAGTGTCTGTTTCAAACTTGATGTGAGGCAATGTGGCACAAATCTCATGCC 840
DB      900 TATCAGGAGTGTCTGTTTCAAACTTGATGTGAGGCAATGTGGCACAAATCTCATGCC 959
QY      841 AGCTCATTAACAGATGAGAACAGCAGTTTATCTACTTAAGACAGATGATGTAGAA 900
DB      960 AGCTCATTAACAGATGAGAACAGCAGTTTATCTACTTAAGACAGATGATGTAGAA 1019
QY      901 AAGGCTGAATTCGTATATAAGCAACACCGCTGTACCAAGGAGCCAAATACAGA 960
DB      1020 AAGGCTGAATTCGTATATAAGCAACACCGCTGTACCAAGGAGCCAAATACAGA 1079
QY      961 TGGGCTGGAAGTAAAGAACATGTAATGATAGGCGGATCCAGACAGAAAAAAGGTA 1020
DB      1080 TGGGCTGGAAGTAAAGAACATGTAATGATAGGCGGATCCAGACAGAAAAAAGGTA 1139
QY      1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGGAATAGCAAACTGCCATGC 1080
DB      1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGGAATAGCAAACTGCCATGC 1199
QY      1081 TCAGAGAACTCTGAGATGATGAAAGATGTCCTGTGATACCTAAATACAGCATTCAG 1140
DB      1200 TCAGAGAACTCTGAGATGATGAAAGATGTCCTGTGATACCTAAATACAGCATTCAG 1259
QY      1141 AAAGTTAATGATGTTTTCAGAAAGTGAATGTTAGGTTCTGTGATGATCATGAT 1200
DB      1260 AAAGTTAATGATGTTTTCAGAAAGTGAATGTTAGGTTCTGTGATGATCATGAT 1319
QY      1201 GGGGAGCTCTAAATCAAAATGCCAAAGTACGTGATGTATGACAGTTCTAAATGAGTAT 1260
DB      1320 GGGGAGCTCTAAATCAAAATGCCAAAGTACGTGATGTATGACAGTTCTAAATGAGTAT 1379
QY      1261 GAATATTTCTGTTCTTCAGAAATATGACTTCTGGCCAGTGATGCTCATGAGGCTTTA 1320
DB      1380 GAATATTTCTGTTCTTCAGAAATATGACTTCTGGCCAGTGATGCTCATGAGGCTTTA 1439
QY      1321 ATATGTAAAGTGAAGAGTTTCACTCCAAATCAGTAGAGATATATTGAAGACAAATA 1380
DB      1440 ATATGTAAAGTGAAGAGTTTCACTCCAAATCAGTAGAGATATATTGAAGACAAATA 1499
QY      1381 TTTGGGAAAACTATGGAAGAGCAAGCTCCCACTTAAGCCATGTATGAAAT 1440
DB      1500 TTTGGGAAAACTATGGAAGAGCAAGCTCCCACTTAAGCCATGTATGAAAT 1559
QY      1441 CTAAATTAAGGACATTTGTTACTGAGCCACATATATCAAGAGGTGCCACAAAT 1500
DB      1560 CTAAATTAAGGACATTTGTTACTGAGCCACATATATCAAGAGGTGCCACAAAT 1619
QY      1501 AAATTAAGGCTTAAAGAGAGACCTACATCAGGCTTCATCTGAGGATTTTATCAAGAAA 1560
DB      1620 AAATTAAGGCTTAAAGAGAGACCTACATCAGGCTTCATCTGAGGATTTTATCAAGAAA 1679
QY      1561 GCAGATTTGGCAGTTCAAAAAGACTCCTGAAATGATTAATCAGGAACTAACAGGAG 1620
DB      1680 GCAGATTTGGCAGTTCAAAAAGACTCCTGAAATGATTAATCAGGAACTAACAGGAG 1739

```

QY	1621	CAGAAATGGTCAAGTCATGAAATATTTATCTATAGTGGTCTATGGAATTAACAAAGGTCAT	1680
Db	1740	CAGAAATGGTCAAGTCATGAAATATTTATCTATAGTGGTCTATGGAATTAACAAAGGTCAT	1799
QY	1681	TCTATTTCAGAAATGAGAAAAATCCATACCCATAGAAATACATCGAAAAAGATCTGCTTC	1740
Db	1800	TCTATTTCAGAAATGAGAAAAATCCATACCCATAGAAATACATCGAAAAAGATCTGCTTC	1859
QY	1741	AAAACGAAAGCTGAACCTATTAAGCAGCATTAAGCAATATGGAATTCGAATTAATATTC	1800
Db	1860	AAAACGAAAGCTGAACCTATTAAGCAGCATTAAGCAATATGGAATTCGAATTAATATTC	1919
QY	1801	CACATTTCAAAAAGCCCTTAAGAAATAGGCTGAGGAGAGAGTCTCTACACAGCATATTC	1860
Db	1920	CACATTTCAAAAAGCCCTTAAGAAATAGGCTGAGGAGAGAGTCTCTACACAGCATATTC	1979
QY	1861	CATGGCGCTTGAACCTAGTAGTCAGTGAATCTAAGCCACCTAATATGACTGAATTCGAA	1920
Db	1980	CATGGCGCTTGAACCTAGTAGTCAGTGAATCTAAGCCACCTAATATGACTGAATTCGAA	2039
QY	1921	ATTGATAGTTGTTCTTAGCAGTGAAGAGATTAAGAAAAAAAGTCAACCAAAATCCAGTC	1980
Db	2040	ATTGATAGTTGTTCTTAGCAGTGAAGAGATTAAGAAAAAAAGTCAACCAAAATCCAGTC	2099
QY	1981	AGGCACACACAAANCTCACTATCTGGAAGCTTAAGAAACCTGCACCTGGAGCCAAAGAG	2040
Db	2100	AGGCACACACAAANCTCACTATCTGGAAGCTTAAGAAACCTGCACCTGGAGCCAAAGAG	2159
QY	2041	AGTAACAAGCCAAATGAACAGACACAAGTAAAAAGACATGCAGTGCATCTTCCAGAGCTG	2100
Db	2160	AGTAACAAGCCAAATGAACAGACACAAGTAAAAAGACATGCAGTGCATCTTCCAGAGCTG	2219
QY	2101	AAGTTAACAAATGCACTGCTGTTCTTTTACTAAGTGTTCAAATACCAGTGAACCTTAAAGAA	2160
Db	2220	AAGTTAACAAATGCACTGCTGTTCTTTTACTAAGTGTTCAAATACCAGTGAACCTTAAAGAA	2279
QY	2161	TTTGTGCAATCCTAGGCTTCCAAAGAGAAACAAAAAGAAAGTGAAGAACTGAAGAACTGTAAGTG	2220
Db	2280	TTTGTGCAATCCTAGGCTTCCAAAGAGAGAAAGAAAGAAAGTGAAGAACTGAAGAACTGTAAGTG	2339
QY	2221	TCTAATAATGCTGAAGACCCCAAGATCTCATGTTAATGAGGAAAGGGTTTTCCAACT	2280
Db	2340	TCTAATAATGCTGAAGACCCCAAGATCTCATGTTAATGAGGAAAGGGTTTTCCAACT	2399
QY	2281	GAAAGATCTGTAAGAGATGACGATATTTCAATTGCTACCTGCTACTGATTAATGACATCAG	2340
Db	2400	GAAAGATCTGTAAGAGATGACGATATTTCAATTGCTACCTGCTACTGATTAATGACATCAG	2459
QY	2341	GAAAGTATCTGTTACTAGGAGTTAGACATCTAGGGAAGGCAAAAAACAGAACCAATATTA	2400
Db	2460	GAAAGTATCTGTTACTAGGAGTTAGACATCTAGGGAAGGCAAAAAACAGAACCAATATTA	2519
QY	2401	TGTGTGAGTCAGTGTGACGATTTTGAAAAACCCCAAGGACATTAATTCATGTGTTTCCAAA	2460
Db	2520	TGTGTGAGTCAGTGTGACGATTTTGAAAAACCCCAAGGACATTAATTCATGTGTTTCCAAA	2579
QY	2461	GATTAATGAAATGACACAGAAAGCTTTAATGATCCATTTGGGACATGAAAGTTAACACAGT	2520
Db	2580	GATTAATGAAATGACACAGAAAGCTTTAATGATCCATTTGGGACATGAAAGTTAACACAGT	2639
QY	2521	CGGGAAACCAACATGAGAAATGGAAGAAAGTGAACCTGATGCTGATATTTTGACAATATCA	2580
Db	2640	CGGGAAACCAACATGAGAAATGGAAGAAAGTGAACCTGATGCTGATATTTTGACAATATCA	2699
QY	2581	TTCAAGGTTTCAAGCGGCACTCATTTGCTCTGTTTCAATCCAGGAAATTCGAGAGAG	2640
Db	2700	TTCAAGGTTTCAAGCGGCACTCATTTGCTCTGTTTCAATCCAGGAAATTCGAGAGAG	2759
QY	2641	GAAATGTCACATTTCTGCCCCACTCTGGGCTCTTAAAGAAACAAAGTCCAAAAGTCACT	2700
Db	2760	GAAATGTCACATTTCTGCCCCACTCTGGGCTCTTAAAGAAACAAAGTCCAAAAGTCACT	2819

QY	2701	TTTTGATGTGAACAAAGAGAAATCAAGGAAGANTGATCTAATATCAAGCTGTA	2766
Db	2820	TTTTGATGTGAACAAAGAGAAATCAAGGAAGANTGATCTAATATCAAGCTGTA	2879
QY	2761	CAGACAGTTAATATCATCTGACAGGGCTTCTGTTGGTTCAGAAAGATTAAGCCAGTTGAT	2820
Db	2880	CAGACAGTTAATATCATCTGACAGGGCTTCTGTTGGTTCAGAAAGATTAAGCCAGTTGAT	2939
QY	2821	AATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGC	2880
Db	2940	AATGCCAAATGTAGTATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGC	2999
QY	2881	AACGAAATGAGACATCTACTCTCAATTAACAATGACCTTTACAAACCCATATCTGTTA	2940
Db	3000	AACGAAATGAGACATCTACTCTCAATTAACAATGACCTTTACAAACCCATATCTGTTA	3059
QY	2941	CCACACATTTTCCCATCAAGTCATTTGTTAAATCAATATCAAGAAATCTGCTAGAG	3000
Db	3060	CCACACATTTTCCCATCAAGTCATTTGTTAAATCAATATCAAGAAATCTGCTAGAG	3119
QY	3001	GAACACTTTGAGGAACATTCATCTGTCACCTGAAGAGAAATGGGAATGAGAACATTTCA	3060
Db	3120	GAACACTTTGAGGAACATTCATCTGTCACCTGAAGAGAAATGGGAATGAGAACATTTCA	3179
QY	3061	AGTACAGTGAACACATTTAGCCGTTATACATTTAGAGAAATGTTTTAAAGAACACAC	3120
Db	3180	AGTACAGTGAACACATTTAGCCGTTATACATTTAGAGAAATGTTTTAAAGAACACAC	3239
QY	3121	TCAACCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGGGGCTCCAGTATTATGAA	3180
Db	3240	TCAACCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGGGGCTCCAGTATTATGAA	3299
QY	3181	ATAGGTTCCAGTGATGAACATTCACAGACATAGGTAGAAACAGAGGCCAAATTTG	3240
Db	3300	ATAGGTTCCAGTGATGAACATTCACAGACATAGGTAGAAACAGAGGCCAAATTTG	3359
QY	3241	AATGTTATGCTAGATTTAGGGGTTTTGGAACCTGAGGCTTTAAACAAGCTTCCTCGGA	3300
Db	3360	AATGTTATGCTAGATTTAGGGGTTTTGGAACCTGAGGCTTTAAACAAGCTTCCTCGGA	3419
QY	3301	AGTAATTTTAAGCATCTCGAATAATAAAAAGCAGAAATATGAAGAATAGTTCAACACTGTT	3360
Db	3420	AGTAATTTTAAGCATCTCGAATAATAAAAAGCAGAAATATGAAGAATAGTTCAACACTGTT	3479
QY	3361	AATACAGATTTCTCTCCATATCTGATTTCAAGTAACTAGAACACGCTATGGGAATAGT	3359
Db	3421	CATGATCTCAGGTTTGTCTGACGACACTGATGACTGTTTATGATGATGGTGAATAAG	3480
QY	3340	CATGATCTCAGGTTTGTCTGACGACACTGATGACTGTTTATGATGATGGTGAATAAG	3599
Db	3481	GAAATATCTAGTTTGTGTAATAATACATTTAAGGAAGCTGCTGTTTTAGCAAAAGC	3540
QY	3420	GAAATATCTAGTTTGTGTAATAATACATTTAAGGAAGCTGCTGTTTTAGCAAAAGC	3600
Db	3541	GTCCGAAAGAGAGACTTAGCAGAGAGTCTTAGCCCTTTACCCATATACATTTGGCTCAG	3660
QY	3460	GTCCGAAAGAGAGACTTAGCAGAGAGTCTTAGCCCTTTACCCATATACATTTGGCTCAG	3719
Db	3601	GGTTACGAGAGAGGGGCCAAGAATTTAGAGTCTCAGAGAGAACTTATCTAGAGGAT	3660
QY	3720	GGTTACGAGAGAGGGGCCAAGAATTTAGAGTCTCAGAGAGAACTTATCTAGAGGAT	3779
Db	3661	GAAAGGCTTCCCTGTTCCAACTGTTTATTTGGTAAAGTAAACATATATCTTCAG	3720
QY	3780	GAAAGGCTTCCCTGTTCCAACTGTTTATTTGGTAAAGTAAACATATATCTTCAG	3839
Db	3721	TCTACTAGGCATAGCACCGTTGCTACACGAGTGTGCTTAAGAACACAGAGAGAAATTTA	3780
QY	3840	TCTACTAGGCATAGCACCGTTGCTACACGAGTGTGCTTAAGAACACAGAGAGAAATTTA	3899
Db	3781	TTATCATTTGAAGATAGCTTAAATGACTGCAGTAAACAGGATATATTTGGCAAGGACATCT	3840

|||||  
3900 TTATCTTGAAGAAATGCTTAAATGACTCAGTAACAGAGTAATATGGCAAGGATCT 3959  
3841 CAGGAACATCACCTTAGAGGAAACAAATGTCGTAGCTGTTTCTTCCACAGTGC 3900  
3960 CAGGAACATCACCTTAGAGGAAACAAATGTCGTAGCTGTTTCTTCCACAGTGC 4019  
3901 AGTAATTTGGAAGACTTGACTGCAATACAAACACCAGAGTCTTCTTGATGGTTC 3960  
4020 AGTAATTTGGAAGACTTGACTGCAATACAAACACCAGAGTCTTCTTGATGGTTC 4079  
3961 TCCAAACAAATGAGCATCAGTCTGAAAGCCAGAGGATGGTCTGAGTGAACAAGATTG 4020  
4080 TCCAAACAAATGAGCATCAGTCTGAAAGCCAGAGGATGGTCTGAGTGAACAAGATTG 4139  
4021 GTTTGATGATGAAGAAGAGGAGGAGGCTTGGAAGAAATATCAAGAGCAAGAAC 4080  
4140 GTTTGATGATGAAGAAGAGGAGGAGGCTTGGAAGAAATATCAAGAGCAAGAAC 4199  
4081 ATGGATTCAAACTTAGGTAGAGCAGCATCTGGGTGTGAGAGTGAACAAGCCTCTGAA 4140  
4200 ATGGATTCAAACTTAGGTAGAGCAGCATCTGGGTGTGAGAGTGAACAAGCCTCTGAA 4259  
4141 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACACTCAGCAGAGGATACCATG 4200  
4260 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACACTCAGCAGAGGATACCATG 4319  
4201 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAACAG 4260  
4320 CAACATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAACAG 4379  
4261 CATGGAGCAGCCTTTTAAACAGCTACCTCCATCATAGTACCTCTCTGCTCTGAG 4320  
4380 CATGGAGCAGCCTTTTAAACAGCTACCTCCATCATAGTACCTCTCTGCTCTGAG 4439  
4321 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACACTCAGCAGAGGATACCATG 4380  
4440 GACTGCTCAGGCTATCTCTCAGAGTGAATTTAACACTCAGCAGAGGATACCATG 4499  
4381 AGTAATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAACAG 4440  
4500 AGTAATTAACCTGATTAAGCTCCAGCAGAAATGGCTGAATAGAACCTGTGTAACAG 4559  
4441 GCAGATAGTCTTACAGCAAGTAAATAAAGAACAGAGTGAAGAGTCACTCCCTCTTAA 4500  
4560 GCAGATAGTCTTACAGCAAGTAAATAAAGAACAGAGTGAAGAGTCACTCCCTCTTAA 4619  
4501 TCCCATCTATTAGATGATAGTGTGATGACATGACAGTGTCTGTGGAGTCTTCAAGATGA 4560  
4620 TCCCATCTATTAGATGATAGTGTGATGACATGACAGTGTCTGTGGAGTCTTCAAGATGA 4679  
4561 AACTACCATCTCAAGAGAGGCTCATTAAGTGTGATGAGAGGAGCAACAGCTGGA 4620  
4680 AACTACCATCTCAAGAGAGGCTCATTAAGTGTGATGAGAGGAGCAACAGCTGGA 4739  
4621 GAGTCTGGGCAACAGATTGAAGCAATCTTACTTCCAGAGCAAGATCTAGAGGA 4680  
4740 GAGTCTGGGCAACAGATTGAAGCAATCTTACTTCCAGAGCAAGATCTAGAGGA 4799  
4681 ACCCTTACCTGGAATCTGGAATCAGCTCTTCTCTGATGACCTGATCTGATCTTCT 4740  
4800 ACCCTTACCTGGAATCTGGAATCAGCTCTTCTCTGATGACCTGATCTGATCTTCT 4859  
4741 GAAGACAGAGCCCAAGAGTCACTGCTGTGGCAACATCCATCTTCAACCTCTGATG 4800  
4860 GAAGACAGAGCCCAAGAGTCACTGCTGTGGCAACATCCATCTTCAACCTCTGATG 4919  
4801 AAAGTTCCTCAATTAAGTTGCAAGATCTGCCAGAGTCCAGTCTGCTCATACTACT 4860  
4920 AAAGTTCCTCAATTAAGTTGCAAGATCTGCCAGAGTCCAGTCTGCTCATACTACT 4979  
4861 GATCTGCTGGGTATATGCAATGGAAGAACTGTGAGCAGGAGAGCAGAAATTGACA 4920  
|||||

Db 4980 GATCTGCTGGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGAGCAGAAATTGACA 5039  
QY 4921 GCTTCAACAGAAAGGGTCAACAAAGAAATGCTCATGGTGTCTGGCTGACCCAGAA 4980  
Db 5040 GCTTCAACAGAAAGGGTCAACAAAGAAATGCTCATGGTGTCTGGCTGACCCAGAA 5099  
QY 4981 GAATTTATCTCTGTCAGAGTTGGCAGAGAAACACACATCACTTAATCTTAAT 5040  
Db 5100 GAATTTATCTCTGTCAGAGTTGGCAGAGAAACACACATCACTTAATCTTAAT 5159  
QY 5041 ACTGAAGAGCTACTGATGTTGTTATGAAGACAGATGCTGAGTTGTGTGAGCAGCA 5100  
Db 5160 ACTGAAGAGCTACTGATGTTGTTATGAAGACAGATGCTGAGTTGTGTGAGCAGCA 5219  
QY 5101 CTGAATATTTTCTAGAAATTTGGGAGGAGAAATGGGTGATGATTTCTGGGAGCC 5160  
Db 5220 CTGAATATTTTCTAGAAATTTGGGAGGAGAAATGGGTGATGATTTCTGGGAGCC 5279  
QY 5161 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGTG 5220  
Db 5280 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGTG 5339  
QY 5221 GTCAATGGAAGAACCCAGAGTCCAAAGCCAGACAGAGATCCAGAGCAAGAAATG 5280  
Db 5340 GTCAATGGAAGAACCCAGAGTCCAAAGCCAGAGAGGAAATCCAGAGCAAGAAATG 5399  
QY 5281 TTCAGAGGCTAGAAATCTGTTGCTATGGGCTTACCAACATGCGCCAGATCAAGT 5340  
Db 5400 TTCAGAGGCTAGAAATCTGTTGCTATGGGCTTACCAACATGCGCCAGATCAAGT 5459  
QY 5341 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5400  
Db 5460 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5519  
QY 5401 GGCACAGGTGTCCACCAATTTGTTGTGACACAGAGTCTGACAGAGCAATGTC 5460  
Db 5520 GGCACAGGTGTCCACCAATTTGTTGTGACACAGAGTCTGACAGAGCAATGTC 5579  
QY 5461 TTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5520  
Db 5580 TTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5639  
QY 5521 AGTGTAGCACTTACAGAGTCCAGAGAGTGTGACACCTTACCTGATACCCAGATCC 5580  
Db 5640 AGTGTAGCACTTACAGAGTCCAGAGAGTGTGACACCTTACCTGATACCCAGATCC 5699  
QY 5581 AGCCACTAC 5589  
Db 5700 AGCCACTAC 5708  
  
RESULT 5  
AR070223 5712 bp DNA linear PAT 18-FEB-2000  
LOCUS AR070223  
DEFINITION Sequence 1 from patent US 5891857.  
ACCESSION AR070223  
VERSION AR070223.1 GI:7221111  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 5712)  
AUTHORS Holt,J.T., Jensen,R.A., King,M.-C., Page,D.L., Szabo,C.I.,  
Jettion,T.L., Robinson-Benion,C.L. and Thompson,M.E.  
TITLE Characterized BRCA1 and BRCA2 proteins and screening and  
therapeutic methods based on characterized BRCA1 and BRCA2 proteins  
JOURNAL Patent: US 5891857-A 1 06-APR-1999;  
FEATURES  
source Location/Qualifiers  
BASE COUNT 1956 a 1099 c 1274 g 1383 t  
ORIGIN

Query Match 99.98; Score 5585.8; DB 6; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGATTTATCTGCTTCGCTGAGAGTACAAAATGTCATTAATGCTATGCGAAA 60  
|||||  
120 ATGATTTATCTGCTTCGCTGAGAGTACAAAATGTCATTAATGCTATGCGAAA 179  
61 ATCTTAGAGTGTCCCATCTGTCTGGATGATCAAGAACCTGTCCACAAAGTGTGAC 120  
|||||  
180 ATCTTAGAGTGTCCCATCTGTCTGGATGATCAAGAACCTGTCTCCACAAAGTGTGAC 239  
121 CACATATTTTGCAAAATTTTGCATGCTGAAATCTTCCACAGAGAAAGGCCCTTCACAG 180  
240 CACATATTTTGCAAAATTTTGCATGCTGAAATCTTCCACAGAGAAAGGCCCTTCACAG 299  
181 TGTCTTATGTAAATGATTAACCAAAAGAGCCTCAAGAAAGTACGAGATTACT 240  
|||||  
300 TGTCTTATGTAAATGATTAACCAAAAGAGCCTCAAGAAAGTACGAGATTACT 359  
241 CAACCTTGTGAAGAGCTATTTGAAAATCATTTGTCTTTCACCTTGACACAGTTTGGAG 300  
|||||  
360 CAACCTTGTGAAGAGCTATTTGAAAATCATTTGTCTTTCACCTTGACACAGTTTGGAG 419  
301 TATGCAAAACAGCTATTAATTTTGCAAAAGAAATTAATCTCTGACATCTAAAGAT 360  
|||||  
420 TATGCAAAACAGCTATTAATTTTGCAAAAGAAATTAATCTCTGACATCTAAAGAT 479  
361 GAAGTTTATCATCAAAAGTATGGGCTACAGAAACCTGCAAAAGATTTCTACAGAT 420  
|||||  
480 GAAGTTTATCATCAAAAGTATGGGCTACAGAAACCTGCAAAAGATTTCTACAGAT 539  
421 GAACCCGAAAATCCTTCTGAGAGAAACAGTCTCAGTGTCAAACTCTTAACCTTGA 480  
540 GAACCCGAAAATCCTTCTGAGAGAAACAGTCTCAGTGTCAAACTCTTAACCTTGA 599  
481 ACTGTGAGAACTCTGAGAGCAAGACGGGATACAACTCAAAAGAGCTGTCTACAT 540  
|||||  
600 ACTGTGAGAACTCTGAGAGCAAGACGGGATACAACTCAAAAGAGCTGTCTACAT 659  
541 GAATTTGGATCGATTTCTGGAATACCGTTAATTAAGGCAACTTAATGCGATGGGA 600  
|||||  
660 GAATTTGGATCGATTTCTGGAATACCGTTAATTAAGGCAACTTAATGCGATGGGA 719  
601 GATCAGAATTTGTACAATCAACCCCTCAAGAGAACGAGATGAATCAGTTTGGATCT 660  
720 GATCAGAATTTGTACAATCAACCCCTCAAGAGAACGAGATGAATCAGTTTGGATCT 779  
661 GCAAAAAAGGCTGCTGTGGAATTTTCTGAGACGATGTACAAATACGATCATCAA 720  
|||||  
780 GCAAAAAAGGCTGCTGTGGAATTTTCTGAGACGATGTACAAATACGATCATCAA 839  
721 CCCAGTAATTAATGATTTGAACACCCTGAGAGCGCTGAGAGGATCCAGAAAAG 780  
|||||  
840 CCCAGTAATTAATGATTTGAACACCCTGAGAGCGCTGAGAGGATCCAGAAAAG 899  
781 TATCAGGCTAGTCTGTTTCAAACTTGATGTGAGCCATGTGACAAATACTCATGCC 840  
900 TATCAGGCTAGTCTGTTTCAAACTTGATGTGAGCCATGTGACAAATACTCATGCC 959  
841 AGCTCATTTACGACATGAGAACAGAGTTTACTCACTAAAGACAGATGATAGTA 900  
|||||  
960 AGCTCATTTACGACATGAGAACAGAGTTTACTCACTAAAGACAGATGATAGTA 1019  
901 AAGGCTGAATCTGTAATTAAGAACAGAGCTGCTTGAAGAGGAGCAACATTAACA 960  
|||||  
1020 AAGGCTGAATCTGTAATTAAGAACAGAGCTGCTTGAAGAGGAGCAACATTAACA 1079  
961 TGGGCTGAATGAAGAAACATGTAATGATAGGCGACTCCACAGACAGAAAAAGGTA 1020  
1080 TGGGCTGAATGAAGAAACATGTAATGATAGGCGACTCCACAGACAGAAAAAGGTA 1139  
1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATTAAGAGAAACTGCCATGC 1080

|||||  
1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGAATTAAGCAAAACTGCCATGC 1199  
1081 TCAGAAATCTAGAGATACGAATGCTCTCTGATTAACACTAAATAGACATTCAG 1140  
1200 TCAGAAATCTAGAGATACGAATGCTCTCTGATTAACACTAAATAGACATTCAG 1259  
1141 AAGTTAATGAGTGTCTTCCAGAGTATGAACCTGTAGGTCTGATGATCACTCAT 1200  
1260 AAGTTAATGAGTGTCTTCCAGAGTATGAACCTGTAGGTCTGATGATCACTCAT 1319  
1201 GGGGAGTCTGAATCAAAATGCCAAATAGCTGATGATTTGAGAGCTTCTAAATAGAGTAT 1260  
1320 GGGGAGTCTGAATCAAAATGCCAAATAGCTGATGATTTGAGAGCTTCTAAATAGAGTAT 1379  
1261 GAATATCTGCTTCTTCAGAGAAATTAACCTTACAGGCGACTGATCCATGAGGCTTTA 1320  
1380 GAATATCTGCTTCTTCAGAGAAATTAACCTTACAGGCGACTGATCCATGAGGCTTTA 1439  
1321 ATATGTAAAGTGAAGAGATTCACCAAAATCAGAGAGTAAATTAATGAAGCAAAATA 1380  
1440 ATATGTAAAGTGAAGAGATTCACCAAAATCAGAGAGTAAATTAATGAAGCAAAATA 1499  
1381 TTTGGAAAACCTATTCGAGAGAGCAAGCCTCCCAACTTAAGCCATGTAACTGAATAAT 1440  
1500 TTTGGAAAACCTATTCGAGAGAGCAAGCCTCCCAACTTAAGCCATGTAACTGAATAAT 1559  
1441 CTAATTTATAGAGCAATTTTGTACTAGAGCAAGATTAATACAAAGAGCCTCCCTACAAAT 1500  
1560 CTAATTTATAGAGCAATTTTGTACTAGAGCAAGATTAATACAAAGAGCCTCCCTACAAAT 1619  
1501 AAATTAAGCGTAAAGAGAGACCTACATCAGAGCCTTCACTCTGAGAGATTTTATCAAAATA 1560  
1620 AAATTAAGCGTAAAGAGAGACCTACATCAGAGCCTTCACTCTGAGAGATTTTATCAAGAAA 1679  
1561 GCAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAGACTAACCAACGGAG 1620  
1680 GCAGATTTGGCAGTTCAAAAGACTCTGAAATGATTAATCAGGAGACTAACCAACGGAG 1739  
1621 CAGAAATGTCAGATGATTAATTAATTAATGCTGATGAGATTAATTAATTAATTAATTAAT 1680  
1740 CAGAAATGTCAGATGATTAATTAATTAATGCTGATGAGATTAATTAATTAATTAATTAAT 1799  
1681 TCTATTCAAGATGAAGAAATCTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740  
1800 TCTATTCAAGATGAAGAAATCTTAACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1859  
1741 AAACGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGAGATCGAATTAATTAATTC 1800  
1860 AAACGAAAGCTGAACCTATTAAGCAGAGTATTAAGCAATATGAGATCGAATTAATTAATTC 1919  
1801 CACAATTCAAAAGACCTTAAGAAAGATATGCTGAGAGGAGAGTCTTACACAGGCAATTC 1860  
1920 CACAATTCAAAAGACCTTAAGAAAGATATGCTGAGAGGAGAGTCTTACACAGGCAATTC 1979  
1861 CATGGCTTGAACCTAGTACGTAGAAATTAAGCCACCTAATTTACTGATTAATTCGA 1920  
1980 CATGGCTTGAACCTAGTACGTAGAAATTAAGCCACCTAATTTACTGATTAATTCGA 2039  
1921 ATTGATATGTTTCTAGCAGTGAAGAGATTAAGAAAAAGTACAAACCAATTCGACATC 1980  
2040 ATTGATATGTTTCTAGCAGTGAAGAGATTAAGAAAAAGTACAAACCAATTCGACATC 2099  
1981 AGGCACAGCAAAACCTCAACCTGATGAAGAGTAAACCTGCAACTGAGAGCAAGAG 2040  
2100 AGGCACAGCAAAACCTCAACCTGATGAAGAGTAAACCTGCAACTGAGAGCAAGAG 2159  
2041 AGTAACAAGCCAAATGAACAGACAGATTAAGAAAGATGACAGTATCTTTCCAGAGCTG 2100  
2160 AGTAACAAGCCAAATGAACAGACAGATTAAGAAAGATGACAGTATCTTTCCAGAGCTG 2219  
2101 AAGTTAACAATGCACTGCTTCTTTACTAAGTGTCAAAATACCACTGAATTAAGAA 2160  
|||||

2220 AAGTAAACAATGCACCTGGTCTTTTACTAAGTGTTCAAATACCACTGAACTTAAGAA 2279  
QY TTTGTCAATCTCTACCTTCCAGAGAGAAAAGAGAAACATAGAAACAGTTAAAGTG 2220  
Db TTTGTCAATCTCTACCTTCCAGAGAGAAAAGAGAAACATAGAAACAGTTAAAGTG 2280  
QY TCTAATATGCTGAGAACCCCAAGATCTCATGTTAAGTGGAGAAAGGTTTTCACAACT 2221  
Db TCTAATATGCTGAGAACCCCAAGATCTCATGTTAAGTGGAGAAAGGTTTTCACAACT 2240  
QY GAAAGATCTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2281  
Db GAAAGATCTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2400  
QY GAAAGATCTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2241  
Db GAAAGATCTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2460  
QY TGTGTAGTCACTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2320  
Db TGTGTAGTCACTGTAGAGAGTAGACAGTATTTCACTGTGTACTGTATATGACACTCAG 2320  
QY GATATATGAATATGACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT 2461  
Db GATATATGAATATGACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT 2580  
QY CGGGAACAAGCATAGAAATGSAAGAAAGTGAAGTATGCTCAGTATTTGCGAAGATACA 2521  
Db CGGGAACAAGCATAGAAATGSAAGAAAGTGAAGTATGCTCAGTATTTGCGAAGATACA 2640  
QY TTTCAAGGTTTCAAGGCCAGTCACTTTGCTCCGTTTCAATCCAGGAATGCGAAGAG 2581  
Db TTTCAAGGTTTCAAGGCCAGTCACTTTGCTCCGTTTCAATCCAGGAATGCGAAGAG 2700  
QY GAATGTGCAACATCTCTGCTGCCACTGCGGTCTTAAAGAAACAACTCCAAAAGTCACT 2641  
Db GAATGTGCAACATCTCTGCTGCCACTGCGGTCTTAAAGAAACAACTCCAAAAGTCACT 2760  
QY TTTGAATGTGAACAAAAGAAAGAAATCAAGAAAGATAGTCTAATATCAAGCCCTGTA 2701  
Db TTTGAATGTGAACAAAAGAAAGAAATCAAGAAAGATAGTCTAATATCAAGCCCTGTA 2820  
QY CAGACAGTATATATCAGTGCAGGCTTCTGTGTGTGTCAGAAAGATAAGCCAGTTGAT 2761  
Db CAGACAGTATATATCAGTGCAGGCTTCTGTGTGTGTCAGAAAGATAAGCCAGTTGAT 2880  
QY AATGCCAAATGTATGATCAAAAGAGGCTCTAGGTTTGTCTATCTCAGTTCAGAGGC 2821  
Db AATGCCAAATGTATGATCAAAAGAGGCTCTAGGTTTGTCTATCTCAGTTCAGAGGC 2940  
QY AAGCAAACTGACATCTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCTGATA 2881  
Db AAGCAAACTGACATCTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCTGATA 3000  
QY CCAACACTTTTCCCATCAAGTCACTTTGTTAAACTTAATTAAGAAAAATCTGCTAGAG 2941  
Db CCAACACTTTTCCCATCAAGTCACTTTGTTAAACTTAATTAAGAAAAATCTGCTAGAG 3060  
QY GAAAACTTTTGAAGAACATTAATGTCACCTGAAAGAGAAATGGGAAATGAGAACATTTCA 3001  
Db GAAAACTTTTGAAGAACATTAATGTCACCTGAAAGAGAAATGGGAAATGAGAACATTTCA 3120  
QY AGTACAGTGAACAATTTAGCCCTAATACATTAGAGAAAAATGTTTTTAAGAACCCAGC 3061  
Db AGTACAGTGAACAATTTAGCCCTAATACATTAGAGAAAAATGTTTTTAAGAACCCAGC 3180  
QY TCAAGCAATATTAATGAAGTGTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3121  
Db TCAAGCAATATTAATGAAGTGTCCAGTACTAATGAAGTGGCTCCAGTATTAATGAA 3240  
QY ATAGGTTCCAGTATGAAGAACATTTCAAGCAGAACTAGTGAAGAAAGAGGCCAAAATTG 3181  
Db ATAGGTTCCAGTATGAAGAACATTTCAAGCAGAACTAGTGAAGAAAGAGGCCAAAATTG 3300

QY AATGCTATGCTTATAGTATGAGGGTTTGGCAACCTGAGGCTATAAACAAGTCTTCTGGA 3241  
Db AATGCTATGCTTATAGTATGAGGGTTTGGCAACCTGAGGCTATAAACAAGTCTTCTGGA 3360  
QY AGTAAATTTGAAGCATCTCTGAAGTAAAAAGCAGAAATATGAAGTATGTTCAAGCTGTT 3301  
Db AGTAAATTTGAAGCATCTCTGAAGTAAAAAGCAGAAATATGAAGTATGTTCAAGCTGTT 3420  
QY AATACAGATTTCTCTCATATCTGATTTTCAGATTAACCTAAGACGCTATGGGAAGTACT 3361  
Db AATACAGATTTCTCTCATATCTGATTTTCAGATTAACCTAAGACGCTATGGGAAGTACT 3480  
QY CATGACATCTAGAGTTTGTGAGACACCTGATGACCTGTAGTATGTTGGAATTAAG 3421  
Db CATGACATCTAGAGTTTGTGAGACACCTGATGACCTGTAGTATGTTGGAATTAAG 3540  
QY GAAGATACATGTTTGTCTGAAGAAATGACATTAAGGAAAGTCTGCTTTTACCAAAAGC 3481  
Db GAAGATACATGTTTGTCTGAAGAAATGACATTAAGGAAAGTCTGCTTTTACCAAAAGC 3600  
QY GTCGAGAAAGAGAGCTTACAGAGAGTCTAGCCCTTTCACCATACACATTTGGCTCAG 3541  
Db GTCGAGAAAGAGAGCTTACAGAGAGTCTAGCCCTTTCACCATACACATTTGGCTCAG 3660  
QY GGTTCGGAAGAGGGGCCAAGAAATTAAGTCTCCAGAAAGAACTATCTAGAGAGAT 3601  
Db GGTTCGGAAGAGGGGCCAAGAAATTAAGTCTCCAGAAAGAACTATCTAGAGAGAT 3720  
QY GAAGAGCTTCCCTGCTTCCAAACATTTGTATTTGTGAAGTAAACAATATACCTTCTCAG 3661  
Db GAAGAGCTTCCCTGCTTCCAAACATTTGTATTTGTGAAGTAAACAATATACCTTCTCAG 3780  
QY TCTACTATGAGCATATGACACCTGTTGCTACCGAGTGTCTTAAGAACACAGAGAGAAATTA 3721  
Db TCTACTATGAGCATATGACACCTGTTGCTACCGAGTGTCTTAAGAACACAGAGAGAAATTA 3840  
QY TTAATCATTTGAAGATATACCTTAATATGACTGACAGTAAATATTTGGCAAGGCACTCT 3781  
Db TTAATCATTTGAAGATATACCTTAATATGACTGACAGTAAATATTTGGCAAGGCACTCT 3900  
QY CAGGAACATCACCTTAAGTGAAGAAACAAATGTTCTGACTGTGTTCTTACACAGTGC 3841  
Db CAGGAACATCACCTTAAGTGAAGAAACAAATGTTCTGACTGTGTTCTTACACAGTGC 3960  
QY AGTGAATTTGGAAGAACCTTGACTGCAAAATACAAACCCAGAGATCTTCTTGATTTGGTCT 3901  
Db AGTGAATTTGGAAGAACCTTGACTGCAAAATACAAACCCAGAGATCTTCTTGATTTGGTCT 4020  
QY TCCAAACAATTAAGGCACTGACTGGAAGGCCAGGAGTTGTGCTGAGTGAAGGAAATG 3961  
Db TCCAAACAATTAAGGCACTGACTGGAAGGCCAGGAGTTGTGCTGAGTGAAGGAAATG 4080  
QY GTTTCAGATGATGAGAAAGAGAGAGGCTTGAAGAAATTAATCAAGAAAGCAAGC 4021  
Db GTTTCAGATGATGAGAAAGAGAGAGGCTTGAAGAAATTAATCAAGAAAGCAAGC 4140  
QY ATGAGATTTCAACTTTAGTGAAGCAGCATCTGGGTGTGAGAGTAAACACGCTCTGAA 4081  
Db ATGAGATTTCAACTTTAGTGAAGCAGCATCTGGGTGTGAGAGTAAACACGCTCTGAA 4200  
QY GACTGCTCAGAGGCTATCTCTCAGAGTGAAGTATTAACCTGACGAGAGGATACCATG 4141  
Db GACTGCTCAGAGGCTATCTCTCAGAGTGAAGTATTAACCTGACGAGAGGATACCATG 4260  
QY CAACATTAACCTGATTAAGCTCCAGCAGAAATGCTGAAGTATGAGTATGAGAACAG 4201  
Db CAACATTAACCTGATTAAGCTCCAGCAGAAATGCTGAAGTATGAGTATGAGAACAG 4320  
QY CATGGAGCCAGGCTTCTAATCAGCTACCTTCCATATAGTATGACTTGTGCTTGGAG 4261  
Db CATGGAGCCAGGCTTCTAATCAGCTACCTTCCATATAGTATGACTTGTGCTTGGAG 4380



```

OY 4321 GACCTCGGAATCCAGAACACATCAGAAAAAGCAGTATTACTTCAGAGAAAAGT 4380
    |||
Db 4440 GACCTCGGAATCCAGAACACATCAGAAAAAGCAGTATTACTTCAGAGAAAAGT 4499
OY 4381 AGTGAATACCCCTTAAGCCAGATCCAGAGGCCCTTCTGCTGACAGTTTGGAGTCT 4440
    |||
Db 4500 AGTGAATACCCCTTAAGCCAGATCCAGAGGCCCTTCTGCTGACAGTTTGGAGTCT 4559
OY 4441 GCAGATAGTCTCCAGTAAATAAAGAACAGAGATGGAAGATCCCTCTTAAA 4500
    |||
Db 4560 GCAGATAGTCTCCAGTAAATAAAGAACAGAGATGGAAGATCCCTCTTAAA 4619
OY 4501 TGCCCATCATTAATAGATAGTGTACATCAGAGTTGCTGTGGAGCTTTCAGATAGA 4560
    |||
Db 4620 TGCCCATCATTAATAGATAGTGTACATCAGAGTTGCTGTGGAGCTTTCAGATAGA 4679
OY 4561 AACTACCCATCTCAGAGAGAGTCAATTAAGTGTGATGTCAGAGAGCAACAGCTGGA 4620
    |||
Db 4680 AACTACCCATCTCAGAGAGAGTCAATTAAGTGTGATGTCAGAGAGCAACAGCTGGA 4739
OY 4621 GAGTCTGGGCGACACAGATTTGACGAAACATCTTACTTGCAGAGCAAGATCTAGAGGA 4680
    |||
Db 4740 GAGTCTGGGCGACACAGATTTGACGAAACATCTTACTTGCAGAGCAAGATCTAGAGGA 4799
OY 4681 ACCCTTACCTGGAATCTGGAATCAGCTCTCTCTGATGACCTGATCTGATCTTCT 4740
    |||
Db 4800 ACCCTTACCTGGAATCTGGAATCAGCTCTCTCTGATGACCTGATCTGATCTTCTCT 4859
OY 4741 GAAGACAGAGCCCGACAGTCTGCTGTGGACATACCTTCAACCTCGCATG 4800
    |||
Db 4860 GAAGACAGAGCCCGACAGTCTGCTGTGGACATACCTTCAACCTCGCATG 4919
OY 4801 AAAGTTCCCAATTGAAAGTTGACAAATCTGCCAGAGTCCAGCTGCTGCTCATACT 4860
    |||
Db 4920 AAAGTTCCCAATTGAAAGTTGACAAATCTGCCAGAGTCCAGCTGCTGCTCATACT 4979
OY 4861 GATACCTGCTGGATATATGCAATGGAAGAAAGTGTGACGCGAGAGAGCAAGATTGACA 4920
    |||
Db 4980 GATACCTGCTGGATATATGCAATGGAAGAAAGTGTGACGCGAGAGAGCAAGATTGACA 5039
OY 4921 GCTTCACAGAAAGGCTCAACAAAGATGTCATGCTGTGTGCTGGCCGACCCGAGAA 4980
    |||
Db 5040 GCTTCACAGAAAGGCTCAACAAAGATGTCATGCTGTGTGCTGGCCGACCCGAGAA 5099
OY 4981 GAATTTATGCTGCTGACAAAGTTTGCAGAAAACACACATCATTAACTTAATCTAAT 5040
    |||
Db 5100 GAATTTATGCTGCTGACAAAGTTTGCAGAAAACACACATCATTAACTTAATCTAAT 5159
OY 5041 ACTGAAGAGACTACTATGTTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGAGACA 5100
    |||
Db 5160 ACTGAAGAGACTACTATGTTGTTATGAAAACAGATGCTGAGTTGTGTGTAACGAGACA 5219
OY 5101 CTGAATATATTTCTAGAAATGCGGAGAGAAATGGGTAGTATTCTTGGGTGACC 5160
    |||
Db 5220 CTGAATATATTTCTAGAAATGCGGAGAGAAATGGGTAGTATTCTTGGGTGACC 5279
OY 5161 CAGTCTATTTAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATG 5220
    |||
Db 5280 CAGTCTATTTAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATG 5339
OY 5221 GTTCATGGAAGAACACCAAGTCCAAAGCGAGCAAGATCCAGAGCAAGAAAGT 5280
    |||
Db 5340 GTTCATGGAAGAACACCAAGTCCAAAGCGAGCAAGATCCAGAGCAAGAAAGT 5399
OY 5281 TTCAGAGGCTTGAATCTGTTGCTATGGGCTTACCAACATGCCCACAGATCAAGT 5340
    |||
Db 5400 TTCAGAGGCTTGAATCTGTTGCTATGGGCTTACCAACATGCCCACAGATCAAGT 5459
OY 5341 GAATGATGATACAGCTGTGTGCTTGTGTGTAAGAGCTTTTCATCATTCACCTT 5400
    |||
Db 5460 GAATGATGATACAGCTGTGTGCTTGTGTGTAAGAGCTTTTCATCATTCACCTT 5519
OY 5401 GGCACAGGTGTCCACCAATTTGTGTGTGTGACAGCCAGATGCTGACAGAGCAATGGC 5460

```

```

Db 5520 GGCACAGGTGTCCACCAATTTGTGTGTGTGACAGCCAGATGCTTGCAGAGCAATGGC 5579
OY 5461 TTCCATGCATTTGGGAGATGTGTGAGGACACCTGTGTGATACCCGAGAGGGTGTGAC 5520
    |||
Db 5580 TTCCATGCATTTGGGAGATGTGTGAGGACACCTGTGTGATACCCGAGAGGGTGTGAC 5639
OY 5521 AGTGTAGCACTTACAGAGTCCAGAGAGTGTGACACCTTACCTGATACCCAGATCCCCAC 5580
    |||
Db 5640 AGTGTAGCACTTACAGAGTCCAGAGAGTGTGACACCTTACCTGATACCCAGATCCCCAC 5699
OY 5581 AGCCACTAC 5589
    |||
Db 5700 AGCCACTAC 5708

RESULT 6
ARL18507 5712 bp DNA linear PAT 16-MAY-2001
LOCUS ARL18507
DEFINITION Sequence 1 from patent US 6149903.
ACCESSION ARL18507
VERSION ARL18507.1 GI:14100417
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5712)
AUTHORS Holt,J.T., Jensen,R.A., King,M.-C., Page,D.L., Szabo,C.I.,
Jettion,T.L., Robinson-Benion,C.L. and Thompson,M.E.
TITLE Characterized BRCA1 and BRCA2 proteins and screening and
JOURNAL therapeutic methods based on characterized BRCA1 and BRCA2 proteins
FEATURES
source 1. 5712
location/Qualifiers
BASE COUNT 1956 a 1099 c 1274 g 1383 t
ORIGIN
Query Match 99.9%; Score 5585.8; DB 6; Length 5712;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 ATGGATTTATCTGCTCTTCCGTTGAAGAGTACAAATGTCATTAATGCTATGACAGAA 60
    |||
Db 120 ATGGATTTATCTGCTCTTCCGTTGAAGAGTACAAATGTCATTAATGCTATGACAGAA 179
OY 61 ATTGTAGAGTCCCATCTGCTGAGGTTGATCAAGAAACCTGTCCACAAAGTGTAC 120
    |||
Db 180 ATTGTAGAGTCCCATCTGCTGAGGTTGATCAAGAAACCTGTGTCCACAAAGTGTAC 239
OY 121 CACATATTTTGCAAATTTTGCATGCTGGAACCTTCAACAGAGAAAGAGGCTTCACAG 180
    |||
Db 240 CACATATTTTGCAAATTTTGCATGCTGGAACCTTCAACAGAGAAAGAGGCTTCACAG 299
OY 181 TGTCTTTTATGTAAGATGATATACCAAGAGGCTTCAAGAAAGTACAGATTTAGT 240
    |||
Db 300 TGTCTTTTATGTAAGATGATATACCAAGAGGCTTCAAGAAAGTACAGATTTAGT 359
OY 241 CAACCTGTGAAGAGCTATTTGAATTCATTTGCTTTTCAAGCTGACAGAGTTTGGAG 300
    |||
Db 360 CAACCTGTGAAGAGCTATTTGAATTCATTTGCTTTTCAAGCTGACAGAGTTTGGAG 419
OY 301 TATGCAACAGCTATTAATTTTGCAGAAATGAGAAATTAATCTCTCGAATCTTAAAGAT 360
    |||
Db 420 TATGCAACAGCTATTAATTTTGCAGAAATGAGAAATTAATCTCTCGAATCTTAAAGAT 479
OY 361 GAAGTTTCTATATCCAAAGTATGGGCTACAGAAACGTCGCCAAAAGACTTCTACAGAGT 420
    |||
Db 480 GAAGTTTCTATATCCAAAGTATGGGCTACAGAAACGTCGCCAAAAGACTTCTACAGAGT 539
OY 421 GAACCCGAATATCTTCTTGTGAGAGAAACATCTCAATGTCACATCTCTTAACCTTGA 480
    |||
Db 540 GAACCCGAATATCTTCTTGTGAGAGAAACATCTCAATGTCACATCTCTTAACCTTGA 599

```



OY	481	ACTGTGAGAACTCTGAGAGACAAAGCAGCGATACAACTCAAAAGACGTCTGTCTACATT	540
OY	482		
OY	483		
OY	484		
OY	485		
OY	486		
OY	487		
OY	488		
OY	489		
OY	490		
OY	491		
OY	492		
OY	493		
OY	494		
OY	495		
OY	496		
OY	497		
OY	498		
OY	499		
OY	500		
OY	501		
OY	502		
OY	503		
OY	504		
OY	505		
OY	506		
OY	507		
OY	508		
OY	509		
OY	510		
OY	511		
OY	512		
OY	513		
OY	514		
OY	515		
OY	516		
OY	517		
OY	518		
OY	519		
OY	520		
OY	521		
OY	522		
OY	523		
OY	524		
OY	525		
OY	526		
OY	527		
OY	528		
OY	529		
OY	530		
OY	531		
OY	532		
OY	533		
OY	534		
OY	535		
OY	536		
OY	537		
OY	538		
OY	539		
OY	540		
OY	541	GAATTGGGATCTGATTCCTCTGGAAGATACCGCTTAATAAGCAACCTTAATTGACAGTGGGA	600
OY	542		
OY	543		
OY	544		
OY	545		
OY	546		
OY	547		
OY	548		
OY	549		
OY	550		
OY	551		
OY	552		
OY	553		
OY	554		
OY	555		
OY	556		
OY	557		
OY	558		
OY	559		
OY	560		
OY	561		
OY	562		
OY	563		
OY	564		
OY	565		
OY	566	GAATTGGGATCTGATTCCTCTGGAAGATACCGCTTAATAAGCAACCTTAATTGACAGTGGGA	719
OY	567		
OY	568		
OY	569		
OY	570		
OY	571		
OY	572		
OY	573		
OY	574		
OY	575		
OY	576		
OY	577		
OY	578		
OY	579		
OY	580		
OY	581		
OY	582		
OY	583		
OY	584		
OY	585		
OY	586		
OY	587		
OY	588		
OY	589		
OY	590		
OY	591		
OY	592		
OY	593		
OY	594		
OY	595		
OY	596		
OY	597		
OY	598		
OY	599		
OY	600		
OY	601	GATCAGAAATGTGTACAAATCACCCCTCAAGAACACAGGAGATGAATACAGTTGGATTCT	660
OY	602		
OY	603		
OY	604		
OY	605		
OY	606		
OY	607		
OY	608		
OY	609		
OY	610		
OY	611		
OY	612		
OY	613		
OY	614		
OY	615		
OY	616		
OY	617		

QY	1561	GCAGATTTGGCAGTTCAAAAGACCTCCTGAATGATAAATTCAGGGAACTACCAAAACGGAG	1620
Db	1680	GCAGATTTTGGCAGTTCAAAAGACCTCCTGAATGATAAATTCAGGGAACTACCAAAACGGAG	1739
QY	1621	CAGATGGTCAAGTATGATTAATTTCTAATAGTGGTCTCATGGAATTAATAACAAAGGTGAT	1680
Db	1740	CAGATGGTCAAGTATGATTAATTTCTAATAGTGGTCTCATGGAATTAATAACAAAGGTGAT	1799
QY	1681	TCTATTGCAAGATGAGAAAAAATCTCAACCAATAGATCTACTCGAAAAAGAAATTCGCTTTC	1740
Db	1800	TCTATTGCAAGATGAGAAAAAATCTCAACCAATAGATCTACTCGAAAAAGAAATTCGCTTTC	1859
QY	1741	AAAACGAAAGCTGAACTATATTAAGCAGCATATTAAGCAATATGAACTCGAATTAATATTC	1800
Db	1860	AAAACGAAAGCTGAACTATATTAAGCAGCATATTAAGCAATATGAACTCGAATTAATATTC	1919
QY	1801	CACATTTCAAAAGCACCCTAAAAAGAAATGGCTGGAGAGAGCTCTCACCAGGCATATT	1860
Db	1920	CACATTTCAAAAGCACCCTAAAAAGAAATGGCTGGAGAGAGCTCTCACCAGGCATATT	1979
QY	1861	CATGGCTTTGAAGTATAGTACATAGTAAGAAATTAAGCCACCCTAATTGTACTGAATTCGA	1920
Db	1980	CATGGCTTTGAAGTATAGTACATAGTAAGAAATTAAGCCACCCTAATTGTACTGAATTCGA	2039
QY	1921	ATTGATAGTGTCTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACAAACCAATCCAGTC	1980
Db	2040	ATTGATAGTGTCTTCTAGCAGTGAAGAGATTAAGAAAAAAGTACAAACCAATCCAGTC	2099
QY	1981	AGGCACACAGAAACCTCAACCTCATGGAAGTAAAGAACCTGCACCTGGAGCCAAAGAG	2040
Db	2100	AGGCACACAGAAACCTCAACCTCATGGAAGTAAAGAACCTGCACCTGGAGCCAAAGAG	2159
QY	2041	AGTACAAGCCCAAAATGAACAGACAAAGTAAAGACATGACATGATCTTCCAGAGCTG	2100
Db	2160	AGTACAAGCCCAAAATGAACAGACAAAGTAAAGACATGACATGATCTTCCAGAGCTG	2219
QY	2101	AAGTTAACCAATGCACCTGTGTTCTTTACTAAGTGTTCAANTACCAGTGAACCTTAAGAA	2160
Db	2220	AAGTTAACCAATGCACCTGTGTTCTTTACTAAGTGTTCAANTACCAGTGAACCTTAAGAA	2279
QY	2161	TTTGTCAATCCGACCTTCCAAAGAGAAAGAAAAAGAAACCTAAGAAACGATTAAGTG	2220
Db	2280	TTTGTCAATCCGACCTTCCAAAGAGAAAGAAAAAGAAACCTAAGAAACGATTAAGTG	2339
QY	2221	TCTAATTAATGCTGAAGACCCCAAGAGATTCATGTTAATGAGAAAGGGTTTCCAACT	2280
Db	2340	TCTAATTAATGCTGAAGACCCCAAGAGATTCATGTTAATGAGAAAGGGTTTCCAACT	2399
QY	2281	GAAAGATCTGAGAGAGTAGAGATTAATTCATTGGTACCTGGTACTGATTAATGGCACTAG	2340
Db	2400	GAAAGATCTGAGAGAGTAGAGATTAATTCATTGGTACCTGGTACTGATTAATGGCACTAG	2459
QY	2341	GAAAGTATCTGTTACTGTGAAGTTAGCACTTAGGGAAGGCAAAAAACAGAACCAATATA	2400
Db	2460	GAAAGTATCTGTTACTGTGAAGTTAGCACTTAGGGAAGGCAAAAAACAGAACCAATATA	2519
QY	2401	TGTGTGATCAGTGTGACGACATTTTGAAGAAACCCCAAGGAGCTAATTCATGTGTTTCCAA	2460
Db	2520	TGTGTGATCAGTGTGACGACATTTTGAAGAAACCCCAAGGAGCTAATTCATGTGTTTCCAA	2579
QY	2461	GATAATFAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGAGACATGAAGTTAACACAGT	2520
Db	2580	GATAATFAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGAGACATGAAGTTAACACAGT	2639
QY	2521	CGGGAACAACACATRAGAATATGGAAGAAAGTAACTTGATGCTCAGTATTTTGCAAAATACA	2580
Db	2640	CGGGAACAACACATRAGAATATGGAAGAAAGTAACTTGATGCTCAGTATTTTGCAAAATACA	2699
QY	2581	TTTCAAGGTTTAAAGCGGCATCTATTTGCTGTGTTTCAATCCAGAAATATGCAGAGAG	2640
Db	2700	TTTCAAGGTTTAAAGCGGCATCTATTTGCTGTGTTTCAATCCAGAAATATGCAGAGAG	2759
QY	2641	GAATGTGCAACATTTCTGTCCCACTGTGGTCTCTTAAAGAAACAAGTCCAAAAGTCACT	2700

```
|||||
Db 2760 GAATGTGCAACATTCCTGCCCCACTCTGGTCTCTTAAGAAACAAAGTCCAAAAGTCAC 2819
OY 2701 TTGTGATGGAACAAAAGAGAAAATCAAGGAAAGATGAGCTATATCAAGCCGTGA 2760
Db 2820 TTTGATGATGAAACAAAAGAGAAAATCAAGGAAAGATGAGCTATATCAAGCCGTGA 2879
OY 2761 CAGACAGTTAATATCACTGACAGCTTCTGTTGGTTCAGAAAGATTAAGCCAGTTGAT 2820
Db 2880 CAGACAGTTAATATCACTGACAGCTTCTGTTGGTTCAGAAAGATTAAGCCAGTTGAT 2939
OY 2821 AATGCCAATATGATATCAAGAGAGCTTAGGTTTTTGTCTATCATCTCAGTTCAAGGC 2880
Db 2940 AATGCCAATATGATATCAAGAGAGCTTAGGTTTTTGTCTATCATCTCAGTTCAAGGC 2999
OY 2881 AACGAACTGAGACTATCTCCAAATTAACATGAGCTTTTACAAACCATATCTGATA 2940
Db 3000 AACGAACTGAGACTATCTCCAAATTAACATGAGCTTTTACAAACCATATCTGATA 3059
OY 2941 CCACCACTTTTCCCATCAAGTCATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3000
Db 3060 CCACCACTTTTCCCATCAAGTCATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3119
OY 3001 GAAACTTTGAGAAACATTCATGTCCACTGAAAGAAATGGGAAATGGAACATTTCCA 3060
Db 3120 GAAACTTTGAGAAACATTCATGTCCACTGAAAGAAATGGGAAATGGAACATTTCCA 3179
OY 3061 AGTAGAGTAGACAAATTCAGCTTAATACATTAGAAAAATGTTTTTAAAGAGCCAG 3120
Db 3180 AGTAGAGTAGACAAATTCAGCTTAATACATTAGAAAAATGTTTTTAAAGAGCCAGC 3239
OY 3121 TCACCAATATTTAATGAAGTAGGTCACATCTAATGAAGTGGGCTCCAGTATTAATGA 3180
Db 3240 TCACCAATATTTAATGAAGTAGGTCACATCTAATGAAGTGGGCTCCAGTATTAATGA 3299
OY 3161 ATAGGTTCCAGTAGAAGAAACATTCAGAGAACTGATAGAAACAGAGGCCCAAAATG 3240
Db 3300 ATAGGTTCCAGTAGAAGAAACATTCAGAGAACTGATAGAAACAGAGGCCCAAAATG 3359
OY 3241 AATGCTATGCTTAGATTAGGGTTTTGCAACCTGAGGCTATTAACAAAGTCTTCTGGA 3300
Db 3360 AATGCTATGCTTAGATTAGGGTTTTGCAACCTGAGGCTATTAACAAAGTCTTCTGGA 3419
OY 3301 AGTAATTTGAGCATCTCAAAATTAAGAAACAGAAATATGAGAGAGTACTGAGCTGT 3360
Db 3420 AGTAATTTGAGCATCTCAAAATTAAGAAACAGAAATATGAGAGAGTACTGAGCTGT 3479
OY 3361 AATACAGATTTCTCCATATCTGATTTGAGATTAAGTGAACAGGCTATGGGAAGTAG 3420
Db 3480 AATACAGATTTCTCCATATCTGATTTGAGATTAAGTGAACAGGCTATGGGAAGTAG 3539
OY 3421 CATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTAGATGATGGTGAATTAAG 3480
Db 3540 CATGATCTCAGGTTGTTCTGAGACACCTGATGACCTGTAGATGATGGTGAATTAAG 3599
OY 3461 GAAGTACTAGTTTGTGTAAGAAATGACATTAAGAAAGTTCTGCTGTTTTTACCAAGC 3540
Db 3600 GAAGTACTAGTTTGTGTAAGAAATGACATTAAGAAAGTTCTGCTGTTTTTACCAAGC 3659
OY 3541 GTCCAGAAAGAGAGCTTAGAGAGAGTCTTACGCCCTTACCCATACATTTGGCTCAG 3600
Db 3660 GTCCAGAAAGAGAGCTTAGAGAGAGTCTTACGCCCTTACCCATACATTTGGCTCAG 3719
OY 3601 GGTTCAGAGAGGGGCCAAGAAATTAGAGTCTCTGAGAGAGAACTTATCTAGTGAAG 3660
Db 3720 GGTTCAGAGAGGGGCCAAGAAATTAGAGTCTCTGAGAGAGAACTTATCTAGTGAAG 3779
OY 3661 GAAGAGCTTCCCTGCTCCACACTGTTATTTGTAAGTAACAAATATACCTTCTCAG 3720
Db 3780 GAAGAGCTTCCCTGCTCCACACTGTTATTTGTAAGTAACAAATATACCTTCTCAG 3839
OY 3721 TCTACTAGGCAATGACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGAATTTA 3780
|||||

Db 3840 TCTACTAGGCAATGACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGAATTTA 3899
OY 3781 TTATCATTTAGAAATATGCTTAATATGATGCTGACATAACAGATTAATTTGGCAAGCATCT 3840
Db 3900 TTATCATTTAGAAATATGCTTAATATGATGCTGACATAACAGATTAATTTGGCAAGCATCT 3959
OY 3841 CAGGAACATCACCTTAGTGAGAGAAACAAAATGTTCTGCTAGCTGTTTTTTCACAGTGC 3900
Db 3960 CAGGAACATCACCTTAGTGAGAGAAACAAAATGTTCTGCTAGCTGTTTTTTCACAGTGC 4019
OY 3901 AGTGAATGAGAGACTGACGTAATATCAAAACCCAGAGATCTTCTGATATGTTGTTCT 3960
Db 4020 AGTGAATGAGAGACTGACGTAATATCAAAACCCAGAGATCTTCTGATATGTTGTTCT 4079
OY 3961 TCCAAACAATGAGGATAGTCTGAAAGCCAGGAGTGGTCTGAGAGTGAACAGAAATG 4020
Db 4080 TCCAAACAATGAGGATAGTCTGAAAGCCAGGAGTGGTCTGAGAGTGAACAGAAATG 4139
OY 4021 GTTTCAGATGATGAAGAAAGAGAAACGGGCTTGGAAAGAAATATCAAGAGCCAAAGC 4080
Db 4140 GTTTCAGATGATGAAGAAAGAGAAACGGGCTTGGAAAGAAATATCAAGAGCCAAAGC 4199
OY 4081 ATGGAATCAACTTTAGTGAAGACACATCTGGTGTGAGAGTGAACAAAGGCTCTGAA 4140
Db 4200 ATGGAATCAACTTTAGTGAAGACACATCTGGTGTGAGAGTGAACAAAGGCTCTGAA 4259
OY 4141 GACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAACCACTCAGCAGAGGATACCATG 4200
Db 4260 GACTGCTCAGGGCTATCTCTCAGAGTGACATTTTAACCACTCAGCAGAGGATACCATG 4319
OY 4201 CAACATTAACCTGATAAGCTCCAGCAGAAATGCTGAACATGAACTGTGTTGAACAG 4260
Db 4320 CAACATTAACCTGATAAGCTCCAGCAGAAATGCTGAACATGAACTGTGTTGAACAG 4379
OY 4261 CATGGAGCCAGCTTTCTACAGCTACCTTCCATCAATGAAGTCTTCTGCCCCTGAG 4320
Db 4380 CATGGAGCCAGCTTTCTACAGCTACCTTCCATCAATGAAGTCTTCTGCCCCTGAG 4439
OY 4321 GACTGCGAATTCAGAAACAAAGCAGATCAGAAAAAGCAGTATTAATTCACAGAAAAAGT 4380
Db 4440 GACTGCGAATTCAGAAACAAAGCAGATCAGAAAAAGCAGTATTAATTCACAGAAAAAGT 4499
OY 4381 AGTGAATTAACCTTATTAAGCCAGAAATCCAGAGGCTTTTCTGCTGACAACTTTGAGTGTCT 4440
Db 4500 AGTGAATTAACCTTATTAAGCCAGAAATCCAGAGGCTTTTCTGCTGACAACTTTGAGTGTCT 4559
OY 4441 GCAGATAGTTCTACAGATAAATAAAGAACAGAGAGTGAAGAAAGTCAATCCCTTTCTAAA 4500
Db 4560 GCAGATAGTTCTACAGATAAATAAAGAACAGAGAGTGAAGAAAGTCAATCCCTTTCTAAA 4619
OY 4501 TGCCCATCATTAAGATAGTGGTGAATGACACAGTTCCTGAGAGTCTTCAGAAATGA 4560
Db 4620 TGCCCATCATTAAGATAGTGGTGAATGACACAGTTCCTGAGAGTCTTCAGAAATGA 4679
OY 4561 AACTTAACCATCTCAGAGAGAGCTCATTAAGTGTGTTGATGTGAGAGAACACAGCTGGAA 4620
Db 4680 AACTTAACCATCTCAGAGAGAGCTCATTAAGTGTGTTGATGTGAGAGAACACAGCTGGAA 4739
OY 4621 GAGTGGGGCCACAGATTTGAGGAAACATTTCTGCAAGGCAAGATCTGAGAGGA 4680
Db 4740 GAGTGGGGCCACAGATTTGAGGAAACATTTCTGCAAGGCAAGATCTGAGAGGA 4799
OY 4681 ACCCTTACCTGGAATCTGGAATCAGGCTCTTCTGTGATGAGCCCTGGAATCTGATCTCT 4740
Db 4800 ACCCTTACCTGGAATCTGGAATCAGGCTCTTCTGTGATGAGCCCTGGAATCTGATCTCTCT 4859
OY 4741 GAAGACAGAGCCCAAGAGTCAAGCTGTTGTGCAACATATCAACTCTGCAATG 4800
Db 4860 GAAGACAGAGCCCAAGAGTCAAGCTGTTGTGCAACATATCAACTCTGCAATG 4919
OY 4801 AAGGTTCCCAATTTGAAGTGTGCAAGATCTGCCAGAGTCCAGTGTCTGCTCATACT 4860
Db 4920 AAGGTTCCCAATTTGAAGTGTGCAAGATCTGCCAGAGTCCAGTGTCTGCTCATACT 4979
|||||
```

QY	4861	GATACGTGGGTAAATTCGAATGGGAAGAAAGTGTGAGCAGGAGCAACCCAGAAATTGACA	4920		
Db	4980	GATACGTGGGTAAATTCGAATGGGAAGAAAGTGTGAGCAGGAGCAACCCAGAAATTGACA	5039		
QY	4921	GCTTCAACAGAAAGGCTCAACAAAAGATGTCATGAGTGTGTGGCTGACCCAGAA	4980		
Db	5040	GCTTCAACAGAAAGGCTCAACAAAAGATGTCATGAGTGTGTGGCTGACCCAGAA	5099		
QY	4981	GAATTATGCTCGTGTACAGTTTGCCAGAAAACACACATCACTTAATCTAAATT	5040		
Db	5100	GAATTATGCTCGTGTACAGTTTGCCAGAAAACACACATCACTTAATCTAAATT	5159		
QY	5041	ACCTAAGAGCATCTCAATGTTGTTATGAAAACACATGCTGAGTTTGTGTGAAAGGACA	5100		
Db	5160	ACCTAAGAGCATCTCAATGTTGTTATGAAAACACATGCTGAGTTTGTGTGAAAGGACA	5219		
QY	5101	CTGAATAATTTTCTAGAGAAATTTGCCGGAGGAAGAAATGGGTAGTAACTATTTCTGGGTGACC	5160		
Db	5220	CTGAATAATTTTCTAGAGAAATTTGCCGGAGGAAGAAATGGGTAGTAACTATTTCTGGGTGACC	5279		
QY	5161	CAGTCTATTTAAAGAAAGAAAATGCTGATGACATGATTTTGAAGTCAGAGAGATGTG	5220		
Db	5280	CAGTCTATTTAAAGAAAGAAAATGCTGATGACATGATTTTGAAGTCAGAGAGATGTG	5339		
QY	5221	GTCATATGGAAGAAACCCACCAAGAGTCGCAAGGAGCAAGAGAAATCCACGACAGAAAGATC	5280		
Db	5340	GTCATATGGAAGAAACCCACCAAGAGTCGCAAGGAGCAAGAGAAATCCACGACAGAAAGATC	5399		
QY	5281	TTCAAGGGGGCTAGAAATCTGTTGCTATGAGGCGCTTCACCAACATGCCACAGATCAACTG	5340		
Db	5400	TTCAAGGGGGCTAGAAATCTGTTGCTATGAGGCGCTTCACCAACATGCCACAGATCAACTG	5459		
QY	5341	GAATGATGCTACAGCTGTGTGTGCTTCTGTGTTGAAGGACGTTTCATCATCAACCTT	5400		
Db	5460	GAATGATGCTACAGCTGTGTGTGCTTCTGTGTTGAAGGACGTTTCATCATCAACCTT	5519		
QY	5401	GGCAGAGGTGCCACCCCAATTGTTGTTGTGAGCAGATGCTGTGAGCAGAGACAAATGGC	5460		
Db	5520	GGCAGAGGTGCCACCCCAATTGTTGTTGTGAGCAGATGCTGTGAGCAGAGACAAATGGC	5579		
QY	5461	TTCCATGCAATTGGGCGAGATGTGTGAGGACACTGTGTGACCCGAGAGATGGGTGTGGAC	5520		
Db	5580	TTCCATGCAATTGGGCGAGATGTGTGAGGACACTGTGTGAGACCCGAGAGATGGGTGTGGAC	5639		
QY	5521	AGGTAGACATCTACCAAGTGGCAGAGCTGGACACCTACCTGATACCCCAATCCCCCAG	5580		
Db	5640	AGGTAGACATCTACCAAGTGGCAGAGCTGGACACCTACCTGATACCCCAATCCCCCAG	5699		
QY	5581	AGCCACTAC 5589			
Db	5700	AGCCACTAC 5708			
RESULT 7	ARI125601	5712 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	ARI125601	5712 bp	DNA	linear	PAT 16-MAY-2001
DEFINITION	Sequence 1 from patent US 6177410.				
ACCESSION	ARI125601				
VERSION	ARI125601.1	GI:14111663			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
AUTHORS	Holt,J.T., Jensen,R.A., King,M.-C., Steiner,M.S., Robinson-Benion,C.L. and Thompson,M.E.				
TITLE	Therapeutic methods for prostate cancer				
JOURNAL	Patent: US 6177410-A 1 23-JAN-2001;				
FEATURES	Location/Qualifiers				
source	1..5712				
BASE COUNT	1956 a 1099 c 1274 g 1383 t				

ORIGIN	Query Match.	99.9%;	Score 5585.8;	DB 6;	Length 5712;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 5587;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	ATGATATTATTCGCTCGCGGCTTGACAAACATACCAAAATGTCATTAAAGCTATGACGAAA	60		
Db	120	ATGATATTTATTCGCTCTTCGCGGTGAAAGATACCAAAATGTCATTAAAGCTATGACGAAA	179		
QY	61	ATCTTAGAGTGTCCCATCTGCTGTGAGTGTGATCAAGAAAGCTGTCTCCACAAGAGTGTAC	120		
Db	180	ATCTTAGAGTGTCCCATCTGCTGTGAGTGTGATCAAGAAAGCTGTCTCCACAAGAGTGTAC	239		
QY	121	CACATATTTTGCAAATTTTGCATGCTGAAACCTTCTCAACGAGAAAGAGGCGTTCCACAG	180		
Db	240	CACATATTTTGCAAATTTTGCATGCTGAAACCTTCTCAACGAGAAAGAGGCGTTCCACAG	299		
QY	181	TGTCCTTTATGTAAAGATGATATACCAAAAGAGCTTACAGAAAGTACAGATTTTGT	240		
Db	300	TGTCCTTTATGTAAAGATGATATACCAAAAGAGCTTACAGAAAGTACAGATTTTGT	359		
QY	241	CACCTTGTGTAAGAGCTATTAAGAAATCATTTGTGCTTTTCAGCTTGTACACAGTTTGGAG	300		
Db	360	CACCTTGTGTAAGAGCTATTAAGAAATCATTTGTGCTTTTCAGCTTGTACACAGTTTGGAG	419		
QY	301	TATGCCAAACAGCTATTAATTTTGCACAAAAGGAAATTAACCTCTCTGAACATCTTAAAGAT	360		
Db	420	TATGCCAAACAGCTATTAATTTTGCACAAAAGGAAATTAACCTCTCTGAACATCTTAAAGAT	479		
QY	361	GAACTTCTATCATTCACAAAGTATGCGCTACAGAAACCGTGCACAAAGACTTCTACAGAGT	420		
Db	480	GAACTTCTATCATTCACAAAGTATGCGCTACAGAAACCGTGCACAAAGACTTCTACAGAGT	539		
QY	421	GAACCCGAAATTCCTCTCTGCAGAGAAACCGTCTCATGCTGCACAACTCTCAACCTTGA	480		
Db	540	GAACCCGAAATTCCTCTCTGCAGAGAAACCGTCTCATGCTGCACAACTCTCAACCTTGA	599		
QY	481	ACTGTGAGAACTCTGAGAGACAAACGAGGATATCAACCTCAAAAGAGCTGTCTACATT	540		
Db	600	ACTGTGAGAACTCTGAGAGACAAACGAGGATATCAACCTCAAAAGAGCTGTCTACATT	659		
QY	541	GAAATGGAGATCTGATTTCTTCTGAAGATACGCTTAATAAGGCACTTATTCAGTGTGGGA	600		
Db	660	GAAATGGAGATCTGATTTCTTCTGAAGATACGCTTAATAAGGCACTTATTCAGTGTGGGA	719		
QY	601	GATCAAGAAATGTATCAAAATCACCCCTCAAGAACCGAGGATGTGAATTCAGTTGGATTC	660		
Db	720	GATCAAGAAATGTATCAAAATCACCCCTCAAGAACCGAGGATGTGAATTCAGTTGGATTC	779		
QY	661	GCAAAAAAGCGCTGTGTGAATTTTCTGAGACGAGATTAACAAATTAAGTGAACATCATCAA	720		
Db	780	GCAAAAAAGCGCTGTGTGAATTTTCTGAGACGAGATTAACAAATTAAGTGAACATCATCAA	839		
QY	721	CCACGATTAATATGATTTTGAACACCACTGAGAAAGCGTGCACGCTGAGAGGCAATCCAGAAAAG	780		
Db	840	CCACGATTAATATGATTTTGAACACCACTGAGAAAGCGTGCACGCTGAGAGGCAATCCAGAAAAG	899		
QY	781	TATCAGAGGTATGTTCTGTTTCAAATTTGCAATGTGAGAGCATGTGGCACAATATCATATGCC	840		
Db	900	TATCAGAGGTATGTTCTGTTTCAAATTTGCAATGTGAGAGCATGTGGCACAATATCATATGCC	959		
QY	841	AGCTCATTTACAGCATGAGACAGAGATTTATTAATCTAAAGACAGATTAATGTAGAA	900		
Db	960	AGCTCATTTACAGCATGAGACAGAGATTTATTAATCTAAAGACAGATTAATGTAGAA	1019		
QY	901	AAGGCTGAATTCGTATTAATAAGCAACAGCGTGGCTTAGCAAGAGACCAACATTAACAGA	960		
Db	1020	AAGGCTGAATTCGTATTAATAAGCAACAGCGTGGCTTAGCAAGAGACCAACATTAACAGA	1079		
QY	961	TGGGCTGGAAGTAAAGAAACATGTAATGATATAGCGGACCTCCACAGACAGAAAAAGCTA	1020		
Db	1080	TGGGCTGGAAGTAAAGAAACATGTAATGATATAGCGGACCTCCACAGACAGAAAAAGCTA	1139		



Dp	3300	ATAGTTCACGTGATGAAACATTCACAGACAGTACGTATGAACAGAGGGCCAAATATG	3359
Qy	3241	AATGCTATGCTTAGATTAGGGGTTTTCACACCTGAGGCTCTATAACAAGTCTTCGGA	3300
Dp	3360	AATGCTATGCTTAGATTAGGGGTTTTCACACCTGAGGCTCTATAACAAGTCTTCGGA	3419
Qy	3301	AGTAATTSTAAGCATCCTCGAATATAAAAAGCAAGATATGAAGAAGTGTGACAGCTGT	3360
Dp	3420	AGTAATTSTAAGCATCCTCGAATATAAAAAGCAAGATATGAAGAAGTGTGACAGCTGT	3479
Qy	3361	AATACAGATTTCTCTCCATATCTGATTTACAGATACTTAGAACACCCATGGGAAGTAGT	3420
Dp	3480	AATACAGATTTCTCTCCATATCTGATTTACAGATACTTAGAACACCCATGGGAAGTAGT	3539
Qy	3421	CATGATCTCAGGTTTGTTCTTGAGACACCTGATGACCTGTATGATGATGGTGAATTAAG	3480
Dp	3540	CATGATCTCAGGTTTGTTCTTGAGACACCTGATGACCTGTATGATGATGGTGAATTAAG	3599
Qy	3481	GAGATACTAGTTTGTCTGAATAATGACATTAAGAAGAAATCTGCTGTTTATGACAAAAG	3540
Dp	3600	GAGATACTAGTTTGTCTGAATAATGACATTAAGAAGAAATCTGCTGTTTATGACAAAAG	3659
Qy	3541	GTCCAGAAAGAGAGCTTAGCAGAGATCTTAGCCCTTTCACCCATPACATATTTGGCTAG	3600
Dp	3660	GTCCAGAAAGAGAGCTTAGCAGAGATCTTAGCCCTTTCACCCATPACATATTTGGCTAG	3719
Qy	3601	GATTACCGAAGAGGGGCCAAGAAATTTAGATCTCAGAGAGAACTTATCTAGAGAGAT	3660
Dp	3720	GATTACCGAAGAGGGGCCAAGAAATTTAGATCTCAGAGAGAACTTATCTAGAGAGAT	3779
Qy	3661	GAAAGAGCTTCCTGCTTCCAAACACTTGTATTTGGTAAAGTAAACAATATACCTTCTCAG	3720
Dp	3780	GAAAGAGCTTCCTGCTTCCAAACACTTGTATTTGGTAAAGTAAACAATATACCTTCTCAG	3839
Qy	3721	TCTACTAGGCCTATACCGGTTGCTACCGAGGCTGTGCTTAGAACACAGAGGAATTTA	3780
Dp	3840	TCTACTAGGCCTATACCGGTTGCTACCGAGGCTGTGCTTAGAACACAGAGGAATTTA	3899
Qy	3781	TTATCATTTGAAGATAGCTTAAATGACTGCAATACACAGATTAATTTGGCAAGGACATCT	3840
Dp	3900	TTATCATTTGAAGATAGCTTAAATGACTGCAATACACAGATTAATTTGGCAAGGACATCT	3959
Qy	3841	CAGGAACATCACCTTAGTGAAGAAACAATATGTTCTGTAGCTGTTTCTTCACAGTGC	3900
Dp	3960	CAGGAACATCACCTTAGTGAAGAAACAATATGTTCTGTAGCTGTTTCTTCACAGTGC	4019
Qy	3901	AGTAATTTGGAAGACTTGACGCAAAATACAAACACCAGGATCCTTTCTTGATTTGGTCT	3960
Dp	4020	AGTAATTTGGAAGACTTGACGCAAAATACAAACACCAGGATCCTTTCTTGATTTGGTCT	4079
Qy	3961	TCCAACAATAGAGGACATCAGTCTGTAAGCCAGGAGTGGTGTAGGTGACAMAGAAATG	4020
Dp	4080	TCCAACAATAGAGGACATCAGTCTGTAAGCCAGGAGTGGTGTAGGTGACAMAGAAATG	4139
Qy	4021	GTTTCAGATGATGAAGAAAGGAAAGGGCTTGGAAATAATATCAAGAAAGACAAGC	4080
Dp	4140	GTTTCAGATGATGAAGAAAGGAAAGGGCTTGGAAATAATATCAAGAAAGACAAGC	4199
Qy	4081	ATGGAATTCAAACTTAGGTGAAGCAGCATCTGGGCTGAGAGTGAAGCAAGGCTCTCGAA	4140
Dp	4200	ATGGAATTCAAACTTAGGTGAAGCAGCATCTGGGCTGAGAGTGAAGCAAGGCTCTCGAA	4259
Qy	4141	GACGCTCAGGGCTATCCTCTCAGAGTACATTTTAACCACTACGACAGAGGATACATG	4200
Dp	4260	GACGCTCAGGGCTATCCTCTCAGAGTACATTTTAACCACTACGACAGAGGATACATG	4319
Qy	4201	CAACATTAACCTGATTAAGCTTCACACAGGAAATGCTCAACTAGAGAAGCTGTGTTGAACAG	4260
Dp	4320	CAACATTAACCTGATTAAGCTTCACACAGGAAATGCTCAACTAGAGAAGCTGTGTTGAACAG	4379
Qy	4261	CATGGAGACCAAGCTTCTAACAGCTACCTTCATCATTAAGTGAAGCTTCTGCCCCTTGAG	4320

Db	4380	CATGGGACCGACGCTTTACAGAGTACCGCTTCCATCAATAGTGACTCTTCCGCTTGAG	4439
OY	4321	GACCTGGGAATCCAGAAACACACATCAGAAAAAGCATTTACTTCACAGAAAGT	4380
Db	4440	GACCTGGCAAAATCCAGAACAAAGACATCAGAAAAAGCATTTACTTCACAGAAAGT	4499
OY	4381	AGTGAATCCCTATTAAAGCAGAAATCCAGAAAGGCTTCTGCTGCAAGTTTGAGTGTCT	4440
Db	4500	AGTGAATCCCTATTAAAGCAGAAATCCAGAAAGGCTTCTGCTGCAAGTTTGAGTGTCT	4599
OY	4441	GCAGATAGTTCTACACAGTAAAAAATAAGAACACAGAGTGAAGATCCCTCTTAAA	4500
Db	4560	GCAGATAGTTCTACACAGTAAAAAATAAGAACACAGAGTGAAGATCCCTCTTAAA	4619
OY	4501	TGCCCATCATTTAGATGATAGTGGTACATGCACAGTTGCTCTGGGAGCTTTCAGATAGA	4560
Db	4620	TGCCCATCATTTAGATGATAGTGGTACATGCACAGTTGCTCTGGGAGCTTTCAGATAGA	4679
OY	4561	AACTAACCATCTCAAGAGAGAGCTCATTAAGTGTTCATATGGAGAGCAACAGCTGGAA	4620
Db	4680	AACTAACCATCTCAAGAGAGAGCTCATTAAGTGTTCATATGGAGAGCAACAGCTGGAA	4739
OY	4621	GAGTCTGGGCCACACGATTTGAGGAAAAACATCTTACTTGCACAGGCAAGATCTTAGAGGGA	4680
Db	4740	GAGTCTGGGCCACACGATTTGAGGAAAAACATCTTACTTGCACAGGCAAGATCTTAGAGGGA	4799
OY	4681	ACCCCTTAACCTGGAGATCTGGAAATAGGCTCTTCATATACCTCGAATCGATCCCTCT	4740
Db	4800	ACCCCTTAACCTGGAGATCTGGAAATAGGCTCTTCATATAGCCTCGAATCGATCCCTCT	4859
OY	4741	GAAGACAGAGCCCGAGAGTCACTGCTGTGGCAACATACCATCTTCAACCTCGCATTG	4800
Db	4860	GAAGACAGAGCCCGAGAGTCACTGCTGTGGCAACATACCATCTTCAACCTCGCATTG	4919
OY	4801	AAAGTTCCCAATTTGAAGAATTGGCAGAAATCTGGCCAGAGTCCAGTGGTCTCATACTACT	4860
Db	4920	AAAGTTCCCAATTTGAAGAATTGGCAGAAATCTGGCCAGAGTCCAGTGGTCTCATACTACT	4979
OY	4861	GATACCTGCTGGGGTAAATGCAATGGAAGAAAGTGTGAGCAGGGGAGAGCCAGAAATTACA	4920
Db	4980	GATACCTGCTGGGGTAAATGCAATGGAAGAAAGTGTGAGCAGGGGAGAGCCAGAAATTACA	5039
OY	4921	GCCTTCAACAGAAAGGGTCAACAAAAAGATGCCATGGTGGTCCGCTGACCCCGAGAA	4980
Db	5040	GCCTTCAACAGAAAGGGTCAACAAAAAGATGCCATGGTGGTCCGCTGACCCCGAGAA	5099
OY	4981	GAATTTATGCGCGGTCAAGTTGGCCAGAAAACACCAACACCTTAACCTAATCTAATT	5040
Db	5100	GAATTTATGCGCGGTCAAGTTGGCCAGAAAACACCAACACCTTAACCTAATCTAATT	5159
OY	5041	ACTGAAGAGACTACTCTATGTTGTTATCAAAAACAGATCCTAAGTTTGTGTGAACGGACA	5100
Db	5160	ACTGAAGAGACTACTCTATGTTGTTATCAAAAACAGATCCTAAGTTTGTGTGAACGGACA	5219
OY	5101	CTGGAATATTTTCTAGGAATTCGGGGAGGAAAAAGGTAATTACTTCTGGGTGACC	5160
Db	5220	CTGGAATATTTTCTAGGAATTCGGGGAGGAAAAAGGTAATTACTTCTGGGTGACC	5279
OY	5161	CAGTCTATTTAAAGAAAAAAATCCTAAATGAGCATATTTTAAAGTCACAGAGAGATGTG	5220
Db	5280	CAGTCTATTTAAAGAAAAAAATCCTAAATGAGCATATTTTAAAGTCACAGAGAGATGTG	5339
OY	5221	GTCATATGGAAGAAACCCCAAGTCCAAAGCGAGCAAGAAATCCACAGACAGAAAGATC	5280
Db	5340	GTCATATGGAAGAAACCCCAAGTCCAAAGCGAGCAAGAAATCCACAGACAGAAAGATC	5399
OY	5281	TTTCAGGGGGCTAGAAAATCTGTGCTATGAGGCGCTTCACCAACATGCCACAGATCAACTG	5340
Db	5400	TTTCAGGGGGCTAGAAAATCTGTGCTATGAGGCGCTTCACCAACATGCCACAGATCAACTG	5459
OY	5341	GAATGATGGTACAGCTGTGTGCTCTTGTGTGTAAGAGAGCTTTCATCATTTCAACCTT	5400
Db	5460	GAATGATGGTACAGCTGTGTGCTCTTGTGTGTAAGAGAGCTTTCATCATTTCAACCTT	5519

QY 5401 GGCACAGGTGTCACACCAATTTGGTGTGAGCCAGATGCTTGACAGAGACATGSC 5460  
|||||  
Db 5520 GGCACAGGTGTCACACCAATTTGGTGTGAGCCAGATGCTTGACAGAGACATGSC 5579  
QY 5461 TTCATGCAATTTGGGCAATGTGTGAGCACCCTGTGTGACCCGAGAGTGGTGTGGAC 5520  
|||||  
Db 5580 TTCATGCAATTTGGGCAATGTGTGAGCACCCTGTGTGACCCGAGAGTGGTGTGGAC 5639  
QY 5521 AGTGAGCACTCTACCACTGCGAGAGCTGGACACCTGCTGATACCCAGATCCCCAC 5580  
|||||  
Db 5640 AGTGAGCACTCTACCACTGCGAGAGCTGGACACCTGCTGATACCCAGATCCCCAC 5699  
QY 5581 AGCCACTAC 5589  
|||||  
Db 5700 AGCCACTAC 5708

RESULT 8  
ARI84044  
LOCUS ARI84044 5712 bp DNA Linear PAT 20-APR-2002  
DEFINITION Sequence 47 from patent US 6342483.  
ACCESSION ARI84044  
VERSION ARI84044.1 GI:20228013  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5712)  
AUTHORS Holt,J.T., Jensen,R.A., Page,D.L., Obermiller,P.S.,  
Robinson-Benion,C.L. and Thompson,M.E.  
TITLE Method for detection and treatment of breast cancer  
JOURNAL Patent: US 6342483-A 47 29-JAN-2002;  
FEATURES  
source 1..5712  
location/Qualifiers  
BASE COUNT 1956 a 1099 c 1274 g 1383 t  
Origin

Query Match 99.9%; Score 5585.8; DB 6; Length 5712;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGATTATCTGCTCTTCGCGTTGAAGAGATACAAATGTCATTATGCTATGCGAATA 60  
|||||  
Db 120 ATGAGATTATCTGCTCTTCGCGTTGAAGAGATACAAATGTCATTATGCTATGCGAATA 179  
QY 61 ATCTTAAGTGTCCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 120  
|||||  
Db 180 ATCTTAAGTGTCCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 239  
QY 121 CACATATTTTGCATAATTTTGCATGCTGAACTTCTCAACAGAAAGGAGGCTTCACAG 180  
|||||  
Db 240 CACATATTTTGCATAATTTTGCATGCTGAACTTCTCAACAGAAAGGAGGCTTCACAG 299  
QY 181 TGTCTTTATGTAGAATGATATTAACCAAGAGAGCTTCAAGAAAGTACAGATTTAGT 240  
|||||  
Db 300 TGTCTTTATGTAGAATGATATTAACCAAGAGAGCTTCAAGAAAGTACAGATTTAGT 359  
QY 241 CAACCTTTGAGAGCTATTGAAATCATTTTGTCTTTTCACTTGCAGTTGACACAGTTTGGAG 300  
|||||  
Db 360 CAACCTTTGAGAGCTATTGAAATCATTTTGTCTTTTCACTTGCAGTTGACACAGTTTGGAG 419  
QY 301 TATGCAAAACAGCTTAATTTTGCAAAAAGGAAATTAACCTCTCTGAAACATCTAAAGAT 360  
|||||  
Db 420 TATGCAAAACAGCTTAATTTTGCAAAAAGGAAATTAACCTCTCTGAAACATCTAAAGAT 479  
QY 361 GAAGTTTCTATCATCAAAAGTATGGCTACAGAAACGCTGCAAAAGACTTCTACAGAGT 420  
|||||  
Db 480 GAAGTTTCTATCATCAAAAGTATGGCTACAGAAACGCTGCAAAAGACTTCTACAGAGT 539  
QY 421 GAACCGAAATCTCTCTTGACAGAAACAGTCTCAGTGTCAACTCTTAACCTTGA 480  
|||||

Db 540 GAACCGAAATCTCTCTTGACAGAAACAGTCTCAGTGTCAACTCTTAACCTTGA 599  
QY 481 ACTGTGAGACCTGAGAGCAAGACGGATTAACCTCAAAAAGAGCTGTGCTCATTT 540  
|||||  
Db 600 ACTGTGAGACCTGAGAGCAAGACGGATTAACCTCAAAAAGAGCTGTGCTCATTT 659  
QY 541 GAATTTGGATCTGATTTCTTCTGAAGATACCGTTAATAAGCAACTTAATTCAGTGTGGGA 600  
|||||  
Db 660 GAATTTGGATCTGATTTCTTCTGAAGATACCGTTAATAAGCAACTTAATTCAGTGTGGGA 719  
QY 601 GATCAAGATTTGTTACAAATCACCCTCAGAGAACCGAGATGAATCACTTTGATTTCT 660  
|||||  
Db 720 GATCAAGATTTGTTACAAATCACCCTCAGAGAACCGAGATGAATCACTTTGATTTCT 779  
QY 661 GCAAAAAAGGCTGCTGTGAATTTTCTGAGACGGATTAACCAATTAACATCAATCA 720  
|||||  
Db 780 GCAAAAAAGGCTGCTGTGAATTTTCTGAGACGGATTAACCAATTAACATCAATCA 839  
QY 721 CCCAGTAATTAATGATTTGMAACCCACCTGAGAACGGTGCAGCTGAGAGGATCCAGAAAG 780  
|||||  
Db 840 CCCAGTAATTAATGATTTGMAACCCACCTGAGAACGGTGCAGCTGAGAGGATCCAGAAAG 899  
QY 781 TATCAGGATGTTCTGTTTCAACTTCGATGTGAGGACCATGTGGCACAAATTAATCTCATGCC 840  
|||||  
Db 900 TATCAGGATGTTCTGTTTCAACTTCGATGTGAGGACCATGTGGCACAAATTAATCTCATGCC 959  
QY 841 AGCTCATTAACAGCATGAGAACAGAGTTTATCTCACTAAACAGAAATGATGTAGAA 900  
|||||  
Db 960 AGCTCATTAACAGCATGAGAACAGAGTTTATCTCACTAAACAGAAATGATGTAGAA 1019  
QY 901 AAGGCTGAATTTCTGTATTAAGCAACACCTGCGCTTACAGAGAGCCCAACATTAACA 960  
|||||  
Db 1020 AAGGCTGAATTTCTGTATTAAGCAACACCTGCGCTTACAGAGAGCCCAACATTAACA 1079  
QY 961 TGGGCTGAGATGAGAGAACATGTATGATGATGCGGACTCCGACACAGAAAAAGGTA 1020  
|||||  
Db 1080 TGGGCTGAGATGAGAGAACATGTATGATGATGCGGACTCCGACACAGAAAAAGGTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATGAGAGAAATGCGCATGC 1080  
|||||  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAAATGAGAGAAATGCGCATGC 1199  
QY 1081 TCAGAGATCTAGAGATCTAGAGATGTTCTTGGATTAACACTTAATATAGACATTTGAG 1140  
|||||  
Db 1200 TCAGAGATCTAGAGATCTAGAGATGTTCTTGGATTAACACTTAATATAGACATTTGAG 1259  
QY 1141 AAAGTTATAGAGTGTTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
|||||  
Db 1260 AAAGTTATAGAGTGTTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1319  
QY 1201 GGGAGTCTGATCAATGCCAAGTAGCTGATGATGATGATGATGATGATGATGATGATGAT 1260  
|||||  
Db 1320 GGGAGTCTGATCAATGCCAAGTAGCTGATGATGATGATGATGATGATGATGATGATGAT 1379  
QY 1361 GAATATTTCTGTTCTTCAGAGAAATAGACTTCTGAGGACGATGATGATGATGATGATGAT 1320  
|||||  
Db 1380 GAATATTTCTGTTCTTCAGAGAAATAGACTTCTGAGGACGATGATGATGATGATGATGAT 1439  
QY 1321 ATATGTAAAGTGAAGAGTTCACCTCAATTCAGTGAAGATATTTGAAGCAAAATA 1380  
|||||  
Db 1440 ATATGTAAAGTGAAGAGTTCACCTCAATTCAGTGAAGATATTTGAAGCAAAATA 1499  
QY 1381 TTTGGGAAACCTATGGAAGAGGAGGCTCCCACTTAAGCATTAAGCATTAAGCATTAAG 1440  
|||||  
Db 1500 TTTGGGAAACCTATGGAAGAGGAGGCTCCCACTTAAGCATTAAGCATTAAGCATTAAG 1559  
QY 1441 CTAATTTATAGAGATTTTCTTCTGAGAGCCACAGATTAATCAAGAGGCTCCCTCAAAAT 1500  
|||||  
Db 1560 CTAATTTATAGAGATTTTCTTCTGAGAGCCACAGATTAATCAAGAGGCTCCCTCAAAAT 1619  
QY 1501 AAATTTAAGGCTAAGAGAGACCTACATCAGGCTTCATCTGAGAGATTTATCAAGAAA 1560  
|||||  
Db 1620 AAATTTAAGGCTAAGAGAGACCTACATCAGGCTTCATCTGAGAGATTTATCAAGAAA 1679



QY	1561	GCAGATTGCGAGTTCAAAAGACTCTGGAATGATTAATAGGCACTAAACCAACGGAG	1620
Db	1680	GCAGATTGGCAGATTCAAAAAGACTCTCGAATGATTAATCAGGGAACCTAACCAACGGAG	1739
QY	1621	CAGAAATGTCAGTGAATGATATTACTAAATAGTGGTCTATGGAATTAACAAAAAGTGAT	1680
Db	1740	CAGAAATGTCAGTGAATGATATTACTAAATAGTGGTCTATGGAATTAACAAAAAGTGAT	1799
QY	1681	TCATTGAGAAATGAGAAAAATCCCTAACCCAAATGAAATCACCTCGAAAAAGAAATGCTTTC	1740
Db	1800	TCATTGAGAAATGAGAAAAATCCCTAACCCAAATGAAATCACCTCGAAAAAGAAATGCTTTC	1859
QY	1741	AAAAAGAAAGCTGAACCTATTAAGCAGCATATAAGCAATATGAACTCGAATTTAAATATC	1800
Db	1860	AAAAAGAAAGCTGAACCTATTAAGCAGCATATAAGCAATATGAACTCGAATTTAAATATC	1919
QY	1801	CACAAATCAAAAGCAGCTTAATAAAGAAATAGGCTGAGAGGAAAGTCTTACCAGCATATT	1860
Db	1920	CACAAATCAAAAGCAGCTTAATAAAGAAATAGGCTGAGAGGAAAGTCTTACCAGCATATT	1979
QY	1861	CATCGCTTGAACAGTAGTACAGTACAGAAATCTTAAGCCACCTAATTTACTGAAATTCGA	1920
Db	1980	CATCGCTTGAACAGTAGTACAGTACAGAAATCTTAAGCCACCTAATTTACTGAAATTCGA	2039
QY	1921	ATTGATAGTTGTTCTTACCATGAAAGAGATAAGCAAAAAAAGTACAACCAATGCCAGTC	1980
Db	2040	ATTGATAGTTGTTCTTACCATGAAAGAGATAAGCAAAAAAAGTACAACCAATGCCAGTC	2099
QY	1981	AGGCACAGCGAAAACCTTACAATCATGGAAGGTAAAGACCTTGCACCTGACGCCAGAG	2040
Db	2100	AGGCACAGCGAAAACCTTACAACCTCATGGAAGGTAAAGACCTTGCACCTGACGCCAGAG	2159
QY	2041	AGTACAAGCCAAATGAAAGACAGCAAGTAAAGACATACAGTGTACTTTCCAGACCTG	2100
Db	2160	AGTACAAGCCAAATGAAAGACAGCAAGTAAAGACATACAGTGTACTTTCCAGACCTG	2219
QY	2101	AAGTTAAACAATGACGACCTGGTCTCTTACTAGTGTCCAATATACCACTGAATTTAAAGAA	2160
Db	2220	AAGTTAAACAATGACGACCTGGTCTCTTACTAGTGTCCAATATACCACTGAATTTAAAGAA	2279
QY	2161	TTTGTCAATCCTAGCCTTCCAGAGAAAGAAAGAGAGAAACTAGAACAGTTAAAGTG	2220
Db	2280	TTTGTCAATCCTAGCCTTCCAGAGAAAGAAAGAGAGAAACTAGAACAGTTAAAGTG	2339
QY	2221	TCATAATATGCTGAAAGACCCCAAAAGCTCATGTTAGTGGAGAAAGGTTTTGCCAACT	2280
Db	2340	TCATAATATGCTGAAAGACCCCAAAAGCTCATGTTAGTGGAGAAAGGTTTTGCCAACT	2399
QY	2281	GAAAGATCTGTAGAGATAGCAGTATTCTATGGTACTGTTATGTGGCAGTACAG	2340
Db	2400	GAAAGATCTGTAGAGATAGCAGTATTCTATGGTACTGTTATGTGGCAGTACAG	2459
QY	2341	GAAAGTATCTGTTACTGGAAGTTAGCAGTCTAGGAGGCCAAAAACAGAACCAATATAA	2400
Db	2460	GAAAGTATCTGTTACTGGAAGTTAGCAGTCTAGGAGGCCAAAAACAGAACCAATATAA	2519
QY	2401	TGTGTAGTGTAGTGTGAGCAATTTGAAAACCCCAAGGAGCAATTAATTCATGGTTGTCCAAA	2460
Db	2520	TGTGTAGTGTAGTGTGAGCAATTTGAAAACCCCAAGGAGCAATTAATTCATGGTTGTCCAAA	2579
QY	2461	GATTAATGAAATGACACAGAGAGCCTTAATATATTCATTTGGACATGAAATTAACACAGT	2520
Db	2580	GATTAATGAAATGACACAGAGAGCCTTAATATATTCATTTGGACATGAAATTAACACAGT	2639
QY	2521	CGGGAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTTGACAGAAATACA	2580
Db	2640	CGGGAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTTGACAGAAATACA	2699
QY	2581	TTTCAAGTTTCAAAAGCCGACAGTCAATTTTGCTGTGTTTCCAATTCAGGAATTCAGAGAG	2640
Db	2700	TTTCAAGTTTCAAAAGCCGACAGTCAATTTTGCTGTGTTTCCAATTCAGGAATTCAGAGAG	2759

QY	2641	GAATGTGCACAACTTCTCTGCCCACCTGGGTCTTAAAGAAACAAATCCAAAGTCACT	2700
Db	2760	GAATGTGCACAACTTCTCTGCCCACCTGGGTCTTAAAGAAACAAATCCAAAGTCACT	2819
QY	2701	TTTGAATGTGAACAAAGAAAGAAATCAAGAAAGATATAGTCTAATATCAAGCTGTA	2760
Db	2820	TTTGAATGTGAACAAAGAAAGAAATCAAGAAAGATATAGTCTAATATCAAGCTGTA	2879
QY	2761	CAGACAGTAAATATCACTCGAGGCTTTCGTGTGGTGTGAGAAAGATTAAGCCAGTTGAT	2820
Db	2880	CAGACAGTAAATATCACTCGAGGCTTTCGTGTGGTGTGAGAAAGATTAAGCCAGTTGAT	2939
QY	2821	AATGCCAAATGTATGATATCAAGAGAGGCTCTAGTTTTGTCTATCATCTCAAGTCAGAGC	2880
Db	2940	AATGCCAAATGTATGATATCAAGAGAGGCTCTAGTTTTGTCTATCATCTCAAGTCAGAGC	2999
QY	2881	AAGGAAACAGGAGCTCTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCGATATA	2940
Db	3000	AAGGAAACAGGAGCTCTTACTCCAAATTAACATGAGACTTTTACAAAACCATATCGATATA	3059
QY	2941	CCACCACCTTTTCCATCAAGTCACTTGTGTTAAACTTAATGTAAGAAAAATCTGCTAGAG	3000
Db	3060	CCACCACCTTTTCCATCAAGTCACTTGTGTTAAACTTAATGTAAGAAAAATCTGCTAGAG	3119
QY	3001	GAAAACCTTTGAGGAGACATTCATGTCACCTGAAAGAGAAATGGGAAATGAGAACATTTCCA	3060
Db	3120	GAAAACCTTTGAGGAGACATTCATGTCACCTGAAAGAGAAATGGGAAATGAGAACATTTCCA	3179
QY	3061	ACTACAGTAGGACAACTTAGCCGTATPACATTAAGAAAAATGTTTTAAAGAACCCAGC	3120
Db	3180	ACTACAGTAGGACAACTTAGCCGTATPACATTAAGAAAAATGTTTTAAAGAACCCAGC	3239
QY	3121	TCACGCAATTTTAATGAAGTAGAGTTCCAGTACTAAATGAAGTGGGCTCCAGTATTAATGAA	3180
Db	3240	TCACGCAATTTTAATGAAGTAGAGTTCCAGTACTAAATGAAGTGGGCTCCAGTATTAATGAA	3299
QY	3181	ATAGGTTCCAGTGATGAAACATTTCAAGCAGAACTAGTAAAGAACAGAGGCCAAATATG	3240
Db	3300	ATAGGTTCCAGTGATGAAACATTTCAAGCAGAACTAGTAAAGAACAGAGGCCAAATATG	3359
QY	3241	AATGCAATCTTAGATTAGGGGTTTTGCAACCGTAGGCTATTAACAAAGTCCTCCAGGA	3300
Db	3360	AATGCAATCTTAGATTAGGGGTTTTGCAACCGTAGGCTATTAACAAAGTCCTCCAGGA	3419
QY	3301	AGTAATTTGTAAACATCTCGAAATTAATAAAGCAAGAATATGAAAGTAGTTCAAGACTGTT	3360
Db	3420	AGTAATTTGTAAACATCTCGAAATTAATAAAGCAAGAATATGAAAGTAGTTCAAGACTGTT	3479
QY	3361	AATACAGATTTCTCTCCATATCTGATTTTCAGATTAAGACAGCCTATGGAGAGTAGT	3420
Db	3480	AATACAGATTTCTCTCCATATCTGATTTTCAGATTAAGACAGCCTATGGAGAGTAGT	3539
QY	3421	CATGCATCTCAGGTTTGTTCTGTGACACCGTAGACTGTTAATAGTATGGTGAATTAAG	3480
Db	3540	CATGCATCTCAGGTTTGTTCTGTGACACCGTAGACTGTTAATAGTATGGTGAATTAAG	3599
QY	3481	GAGATACTAGTTTTGCTGAAATATGACATTAAGGAAAGTTCTGCTGTTTTAGCAAAAAGC	3540
Db	3600	GAGATACTAGTTTTGCTGAAATATGACATTAAGGAAAGTTCTGCTGTTTTAGCAAAAAGC	3659
QY	3541	GTCACGAAAAGGAGAGCTTAGCAGAGAGTCTAGCCCTTTTACCACCATACATTTGGCTCAG	3600
Db	3660	GTCACGAAAAGGAGAGCTTAGCAGAGAGTCTAGCCCTTTTACCACCATACATTTGGCTCAG	3719
QY	3601	GGTTACCCGAAAGAGGGGCCAAAGAAATTAAGAGTCTCAGAAAGAACTTATCTAGGAGAT	3660
Db	3720	GGTTACCCGAAAGAGGGGCCAAAGAAATTAAGAGTCTCAGAAAGAACTTATCTAGGAGAT	3779
QY	3661	GAAAGAGCTTCCCTGCTTCCAAACACTGTTATTTGGTAAAGTAAACATATATACCTTCTCAG	3720
Db	3780	GAAAGAGCTTCCCTGCTTCCAAACACTGTTATTTGGTAAAGTAAACATATATACCTTCTCAG	3839
QY	3721	TCTACTTAGGCATAGCACCGTGTGCTACCGAGTGTCTGTTAAGAACACAGAGAGAAATTTA	3780

|||||  
Db 3840 TCTACTAGCATAGCACCCTGGCTACCGAGTGTCTGTCTAAGAACACAGAGGAATTTA 3899  
OY 3781 TTATCATTAAGAAATAGCTTAAATAGCTGACGTATACCAGGTAATATTTGGCAAGGCATCT 3840  
Db 3900 TTATCATTAAGAAATAGCTTAAATAGCTGACGTATACCAGGTAATATTTGGCAAGGCATCT 3959  
OY 3841 CAGGAACATCACCCTTAGTGAAGAAACAAATGTTGTGCTAGCTTGTCTTCTTCTACAGTGC 3900  
Db 3960 CAGGAACATCACCCTTAGTGAAGAAACAAATGTTGTGCTAGCTTGTCTTCTTCTACAGTGC 4019  
OY 3901 AGTGAATTTGAAGACTTGACTGCAATATCAAAACACCAGAGATCCTTCTTGAATTTGCTCT 3960  
Db 4020 AGTGAATTTGAAGACTTGACTGCAATATCAAAACACCAGAGATCCTTCTTGAATTTGCTCT 4079  
OY 3961 TCCAAACAATAGGATCAGCTGAAAGCCAGGAGTGTGCTGAGTGAACAAGAAATG 4020  
Db 4080 TCCAAACAATAGGATCAGCTGAAAGCCAGGAGTGTGCTGAGTGAACAAGAAATG 4139  
OY 4021 GTTTCAGATGATGAAGAAAGAGAACGGCTTGAAGAAATATCAAGAGAGCAAGAC 4080  
Db 4140 GTTTCAGATGATGAAGAAAGAGAACGGCTTGAAGAAATATCAAGAGAGCAAGAC 4199  
OY 4081 ATGATTTCAAACTTAGGTGAGCAGCAGATCTGGGTGTGAGAGTGAACAAGCCTCTCTGA 4140  
Db 4200 ATGATTTCAAACTTAGGTGAGCAGCAGATCTGGGTGTGAGAGTGAACAAGCCTCTCTGA 4259  
OY 4141 GACTGCTGAGGCTTTCCTCTGAGAGTGACATTTTAAACGCTCAGAGAGGGATACCATG 4200  
Db 4260 GACTGCTGAGGCTTTCCTCTGAGAGTGACATTTTAAACGCTCAGAGAGGGATACCATG 4319  
OY 4201 CAACATTAACCTGATTAAGCTCCAGCAGGAATGGCTGAACCTAGAACCTGTGTAAGACAG 4260  
Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGGAATGGCTGAACCTGTGTAAGACAG 4379  
OY 4261 CATGGGACCGACCTTCTTAACAGCTACCTTCCATCATTAAGTACTCTTCTGCTTGAAG 4320  
Db 4380 CATGGGACCGACCTTCTTAACAGCTACCTTCCATCATTAAGTACTCTTCTGCTTGAAG 4439  
OY 4321 GACCTGCGCAATTCAGAACAAAGCAATCAGAAAAAGCAGTATTACTTACAGAAAAAT 4380  
Db 4440 GACCTGCGCAATTCAGAACAAAGCAATCAGAAAAAGCAGTATTACTTACAGAAAAAT 4499  
OY 4381 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGCTGACAAGTTTGAAGTCT 4440  
Db 4500 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGCTGACAAGTTTGAAGTCT 4559  
OY 4441 GCAGATTAAGCTTACAGTAAAAATTAAGAACCCAGAGGTGAAGAGTCAATCCCTCTTAA 4500  
Db 4560 GCAGATTAAGCTTACAGTAAAAATTAAGAACCCAGAGGTGAAGAGTCAATCCCTCTTAA 4619  
OY 4501 TGCCCATTTAGATAGTGTGATGACATGACAGTGTGCTGAGAGTCTTCAAGATAGA 4560  
Db 4620 TGCCCATTTAGATAGTGTGATGACATGACAGTGTGCTGAGAGTCTTCAAGATAGA 4679  
OY 4561 AACTRACCATCTCAAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGAA 4620  
Db 4680 AACTRACCATCTCAAGAGAGCTCATTAAGTTGTTGATGTGAGAGCAACAGCTGAA 4739  
OY 4621 GAGTGTGGGCAACAGATTTGAGGGAACATCTTCTGCTCCAAAGCAACATCTAAGAGGA 4680  
Db 4740 GAGTGTGGGCAACAGATTTGAGGGAACATCTTCTGCTCCAAAGCAACATCTAAGAGGA 4799  
OY 4681 ACCCTTACCTGGAATCTGGAATCAGCTTCTCTGATGACCTGGAATCTGATCTTCT 4740  
Db 4800 ACCCTTACCTGGAATCTGGAATCAGCTTCTCTGATGACCTGGAATCTGATCTTCT 4859  
OY 4741 GAAGACAGAGCCCGAGAGTCACTGCTGTGGAACATACATCTTCAACCTCTGCAATG 4800  
Db 4860 GAAGACAGAGCCCGAGAGTCACTGCTGTGGAACATACATCTTCAACCTCTGCAATG 4919  
OY 4801 AAAGTCCCAATTGAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTACT 4860  
|||||

Db 4920 AAAGTCCCAATTGAAGTTGCAAAATCTGCCAGAGTCCAGCTGCTCATACTACT 4979  
OY 4861 GATACCTGCTGGGTATATATCATATGGAAGAAATGTGACGAGGAGAAACCCAGAAATGCA 4920  
Db 4980 GATACCTGCTGGGTATATATCATATGGAAGAAATGTGACGAGGAGAAACCCAGAAATGCA 5039  
OY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCCATGCTGTGTGCTGACCCAGAA 4980  
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCCATGCTGTGTGCTGACCCAGAA 5099  
OY 4981 GAATTTATCTGCTGTGACAGTTTCCAGAAACCAACATCATCTTAACTAATTAAT 5040  
Db 5100 GAATTTATCTGCTGTGACAGTTTCCAGAAACCAACATCATCTTAACTAATTAAT 5159  
OY 5041 ACTGAAGACATCACTATGCTGTGTGAAGAAACAGATGCTGAGTGTGGAACGCA 5100  
Db 5160 ACTGAAGACATCACTATGCTGTGTGAAGAAACAGATGCTGAGTGTGGAACGCA 5219  
OY 5101 CTGAATATTTCTAGCAATTCGCGAGAGGAAATGGTAGTTAGCTATTCTGGGTGAC 5160  
Db 5220 CTGAATATTTCTAGCAATTCGCGAGAGGAAATGGTAGTTAGCTATTCTGGGTGAC 5279  
OY 5161 CAGCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAATGACAGAGATGTG 5220  
Db 5280 CAGCTATTTAAAGAAAGAAATGCTGAATGACATGATTTTGAATGACAGAGATGTG 5339  
OY 5221 GTCAATGAGAAAGAACCCAGGCTCAAGGAGGAGCAAGATCCAGAGAGAAATG 5280  
Db 5340 GTCAATGAGAAAGAACCCAGGCTCAAGGAGGAGCAAGATCCAGAGAGAAATG 5399  
OY 5281 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCAACATGCCAGATCAACTG 5340  
Db 5400 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCCCTTACCAACATGCCAGATCAACTG 5459  
OY 5341 GAATGATGTGACAGCTGTGTGCTGTGTGTGTGAAGAGCTTTCATTTACCTT 5400  
Db 5460 GAATGATGTGACAGCTGTGTGCTGTGTGTGTGAAGAGCTTTCATTTACCTT 5519  
OY 5401 GGCACAGGTGTCCACCAATTTGTGTGTGTGACAGAGATGCTGACAGAGAAATG 5460  
Db 5520 GGCACAGGTGTCCACCAATTTGTGTGTGTGACAGAGATGCTGACAGAGAAATG 5579  
OY 5461 TTCATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGTGGGTGTGAC 5520  
Db 5580 TTCATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGTGGGTGTGAC 5639  
OY 5521 AGTGTAGCCTCTACAGAGTCCAGAGAGCTGACACCTTACCTGATACCCCAATCC 5580  
Db 5640 AGTGTAGCCTCTACAGAGTCCAGAGAGCTGACACCTTACCTGATACCCCAATCC 5699  
OY 5581 AGCCACTAC 5589  
Db 5700 AGCCACTAC 5708

RESULT 9  
AR004673  
LOCUS AR004673 5914 bp DNA Linear PAT 04-DEC-1998  
DEFINITION Sequence 1 from patent US 5747282.  
ACCESSION AR004673  
VERSION AR004673.1 GI:3965552  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5914)  
AUTHORS Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
Hershman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.  
and Futreal,P.Andrew.  
170-linked breast and ovarian cancer susceptibility gene  
Patent: US 5747282-A 1 05-MAY-1998;  
FEATURES  
Location/Qualifiers  
source 1..5914



BASE COUNT 2006 a 1156 c 1316 g 1436 t  
ORIGIN /organism="unknown"  
Query Match 99.9%; Score 5585.8; DB 6; Length 5914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5387; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGATTTATCTGCTCTGCGTTGGAAGATCAAAATGCTATATGATGACGAAA 60  
120 ATGATTTATCTGCTCTGCGTTGGAAGATCAAAATGCTATATGATGACGAAA 179  
DB 61 ATCTAGAGTGTCCATCTGCTGAGTTGATCAGAACTGTCCACAAAGTGTGAC 120  
180 ATCTAGAGTGTCCATCTGCTGAGTTGATCAGAACTGTCCACAAAGTGTGAC 229  
QY 121 CACATATTTTGCAATTTTGATGCTGAAACTTCTACACAGAAAGGCGCTTACAG 180  
240 CACATATTTTGCAATTTTGATGCTGAAACTTCTACACAGAAAGGCGCTTACAG 299  
DB 181 TGTCTTTATGTAGATGATATATACCAAAAGAGCGCTACAGAAAGTACGATTTAGT 240  
300 TGTCTTTATGTAGATGATATATACCAAAAGAGCGCTACAGAAAGTACGATTTAGT 359  
QY 241 CAACCTGTGGAAGAGCTATTTGAAATCATTTTGTCTTTTCAAGCTTACACAGGTTTGGAG 300  
360 CAACCTGTGGAAGAGCTATTTGAAATCATTTTGTCTTTTCAAGCTTACACAGGTTTGGAG 419  
DB 301 TATGCAACAGCTATATTTTGGCAAAAAGAAATTAATCTCTCTACACATCTTAAAGAT 360  
420 TATGCAACAGCTATATTTTGGCAAAAAGAAATTAATCTCTCTACACATCTTAAAGAT 479  
QY 361 GAAGTTTCTATCATCCAAAGTATGAGCTCAGAAACCGTGCCAAAGAGCTTACAGAGT 420  
480 GAAGTTTCTATCATCCAAAGTATGAGCTCAGAAACCGTGCCAAAGAGCTTACAGAGT 539  
DB 421 GAACCCGAAATCTCTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 480  
540 GAACCCGAAATCTCTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 599  
QY 481 ACTGTAGAACTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 540  
600 ACTGTAGAACTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 659  
DB 541 GAATGGAGTGTGATCTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 600  
660 GAATGGAGTGTGATCTCTGAGAAACCAAGTCTGAGTCCACCTCTTAACTTGA 719  
QY 601 GATCAAGATTTGTACAAATCAACCCCTCAAGGAACCGAGGATGAATCAGTTTGAATCT 660  
720 GATCAAGATTTGTACAAATCAACCCCTCAAGGAACCGAGGATGAATCAGTTTGAATCT 779  
DB 661 GCAAAAAAGGCTGTTGTAATTTTGTGAGAGGATGTAACTAATCTCAATCTATCA 720  
780 GCAAAAAAGGCTGTTGTAATTTTGTGAGAGGATGTAACTAATCTCAATCTATCA 839  
QY 721 CCCAGTAATATGATTTGAAACCAACCACTGAGAAAGCGTGAGAGGATCGAGAAAG 780  
840 CCCAGTAATATGATTTGAAACCAACCACTGAGAAAGCGTGAGAGGATCGAGAAAG 899  
DB 841 TATCAGGAGTGTCTGTTTCAAACTTGATGTGAGAGGATGTGAGAGGATCGAGAAAG 900  
900 TATCAGGAGTGTCTGTTTCAAACTTGATGTGAGAGGATGTGAGAGGATCGAGAAAG 959  
QY 901 AACCTATTACAGATGAGAAACAGAGTTTATCTACATAAGAGAGATGATGTAGAA 900  
960 AACCTATTACAGATGAGAAACAGAGTTTATCTACATAAGAGAGATGATGTAGAA 1019  
DB 901 AAGGCTGAATTTCTGTATATAAAGCAAAAGAGGCTGTTAGCAAGAGGAGGAGGAGGAG 960  
1020 AAGGCTGAATTTCTGTATATAAAGCAAAAGAGGCTGTTAGCAAGAGGAGGAGGAGGAG 1079  
QY 961 TGGGCTGAAGTGAAGAAACATGTATGTAGGCGGAGCTCCAGGACAGAAAAAAGGTA 1020

DB 1080 TGGGCTGAAGTGAAGAAACATGTATGTAGGCGGAGCTCCAGGACAGAAAAAAGGTA 1139  
QY 1021 GATCTGAATGCTGATCCCTGCTGAGAGAAAGAGTGAATATGAGAGAAACAGGAGGAG 1080  
DB 1140 GATCTGAATGCTGATCCCTGCTGAGAGAAAGAGTGAATATGAGAGAAACAGGAGGAG 1199  
QY 1081 TCAGAGAAATCTGAGATGATGAGAGTGTCTGAGTAAACACTTAAATAGCAGATTCAG 1140  
1200 TCAGAGAAATCTGAGATGATGAGAGTGTCTGAGTAAACACTTAAATAGCAGATTCAG 1259  
DB 1141 AAGGTTAATGAGTGTCTTCCAGAAATGATGAACCTGTAGTGTGATGATCTACATGAT 1200  
1260 AAGGTTAATGAGTGTCTTCCAGAAATGATGAACCTGTAGTGTGATGATCTACATGAT 1319  
QY 1201 GGGGAGTCTGATCAATATGCAAAAGTGTGATGTATTTGGAGCTTCTAAATAGGTAGAT 1260  
1320 GGGGAGTCTGATCAATATGCAAAAGTGTGATGTATTTGGAGCTTCTAAATAGGTAGAT 1379  
DB 1261 GAATATTCTGCTTCTCAGAGAAATAGACTTCTGAGGAGGATGATCTCAGAGGCTTTA 1320  
1380 GAATATTCTGCTTCTCAGAGAAATAGACTTCTGAGGAGGATGATCTCAGAGGCTTTA 1439  
QY 1321 ATATGTAAAGTGAAGAGTCTCACTCAATCAGTAGAGATATATTTGAAGCAAAATA 1380  
1440 ATATGTAAAGTGAAGAGTCTCACTCAATCAGTAGAGATATATTTGAAGCAAAATA 1499  
DB 1381 TTTGGGAAACCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
1500 TTTGGGAAACCTATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1559  
QY 1441 CTATATTATGAGACATTTTGTACTGAGCCACAGATATATCAAGAGGAGGAGGAGGAG 1500  
1560 CTATATTATGAGACATTTTGTACTGAGCCACAGATATATCAAGAGGAGGAGGAGGAGGAG 1619  
DB 1501 AAATTAAGCGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
1620 AAATTAAGCGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1679  
QY 1561 GCAGATTTGGGAGTCTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620  
1680 GCAGATTTGGGAGTCTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1739  
DB 1621 CAGAAATGCTGAGTGTGATTAATTAATAGTGTGATGAGGAGGAGGAGGAGGAGGAGGAG 1680  
1740 CAGAAATGCTGAGTGTGATTAATTAATAGTGTGATGAGGAGGAGGAGGAGGAGGAGGAG 1799  
QY 1741 AAAAGCAAAAGCTGACCTTAATAGCAGCAGATTAAGCAATATGAGGAGGAGGAGGAGGAG 1800  
1860 AAAAGCAAAAGCTGACCTTAATAGCAGCAGATTAAGCAATATGAGGAGGAGGAGGAGGAG 1919  
DB 1801 CACAATTTCAAAAGCAGCTTAAAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860  
1920 CACAATTTCAAAAGCAGCTTAAAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1979  
QY 1861 CATGCGCTTGAAGTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
1980 CATGCGCTTGAAGTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2039  
DB 1921 ATTTGATGATGTTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
2040 ATTTGATGATGTTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2099  
QY 1981 AGGACAGCAGAAACCTTACAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040  
2100 AGGACAGCAGAAACCTTACAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2159  
DB 2041 AGTAAACGCAAAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100

Db 2160 AGTAACAAAGCCAAATGTAACAGACAAAGTAAAAAGACATGACAGCGATCTTTCCAGAGCTG 2219  
OY 2101 AAGTTAACAAATGACACCTGTTCTTTTACTAAGTGTTCAAATACGATGCACTTAAGAA 2160  
Db 2220 AAGTTAACAAATGACACCTGTTCTTTTACTAAGTGTTCAAATACGATGCACTTAAGAA 2279  
OY 2161 TTTGTCAATCCTAGCCCTTCCAGAGAGAAGAAAAAGAGAACTAGAAACAGTTAAAGTG 2220  
Db 2280 TTTGTCAATCCTAGCCCTTCCAGAGAGAAGAAAAAGAGAACTAGAAACAGTTAAAGTG 2339  
OY 2221 TCTAATATGCTGAAGACCCCAAGAAAGATCTCATGTTAAGTGAAGAGGGTTTGCAACT 2280  
Db 2340 TCTAATATGCTGAAGACCCCAAGAAAGATCTCATGTTAAGTGAAGAGGGTTTGCAACT 2339  
OY 2281 GAAAGATCTGTAGAGAGTAGACAGTATTTGATTTGATGCTGCTGATTTATGACACTCAG 2340  
Db 2400 GAAAGATCTGTAGAGAGTAGACAGTATTTGATTTGATGCTGCTGATTTATGACACTCAG 2459  
OY 2341 GAAAGATCTGTAGAGAGTAGACAGTATTTGATTTGATGCTGCTGATTTATGACACTCAG 2400  
Db 2460 GAAAGATCTGTAGAGAGTAGACAGTATTTGATTTGATGCTGCTGATTTATGACACTCAG 2519  
OY 2401 TGTGTGAGTCAGTGTGCAACATTTGAAACCCCAAGGACTAATTTGATGCTGCTGCAAA 2460  
Db 2520 TGTGTGAGTCAGTGTGCAACATTTGAAACCCCAAGGACTAATTTGATGCTGCTGCAAA 2579  
OY 2461 GATTAATAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT 2520  
Db 2580 GATTAATAGAAATGACACAGAAAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT 2639  
OY 2521 CGGGAAACAAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCAAGATCA 2580  
Db 2640 CGGGAAACAAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCAAGATCA 2699  
OY 2581 TTTCAAGTTTCAAAAGCGCAGTCACTTGTCTGTTTCAATTCAGAAATGACAGAAAG 2640  
Db 2700 TTTCAAGTTTCAAAAGCGCAGTCACTTGTCTGTTTCAATTCAGAAATGACAGAAAG 2759  
OY 2641 GAATGTGCAACATCTCTGCCCAGTCTGTGGCTCTTAAGAAACAAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGCAACATCTCTGCCCAGTCTGTGGCTCTTAAGAAACAAAGTCCAAAGTCACT 2819  
OY 2701 TTTGATGTGAAACAAAGAGAAAGAAATCAAGAAAGATGATCTATATTCAGAGCTGTA 2760  
Db 2820 TTTGATGTGAAACAAAGAGAAAGAAATCAAGAAAGATGATCTATATTCAGAGCTGTA 2879  
OY 2761 CAGACAGTTAATATCATCTGCAAGGCTTCTCTGTGGTGGTCAGAAAGATGAAGCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCATCTGCAAGGCTTCTCTGTGGTGGTCAGAAAGATGAAGCAGTTGAT 2939  
OY 2821 AATGCCAAATGTAGTATCAAAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGAGC 2880  
Db 2940 AATGCCAAATGTAGTATCAAAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGAGC 2999  
OY 2881 AACGAAATGAGCTATCTATCTCCAAATAACATGCACTTTTACAAAACCCATATGCTATA 2940  
Db 3000 AACGAAATGAGCTATCTATCTCCAAATAACATGCACTTTTACAAAACCCATATGCTATA 3059  
OY 2941 CCACACATTTTCCCATCAAGTCATTTGTTAAACTAAATGTAAAGAAAAATCGCTAGAG 3000  
Db 3060 CCACACATTTTCCCATCAAGTCATTTGTTAAACTAAATGTAAAGAAAAATCGCTAGAG 3119  
OY 3001 GAAAACTTTGAGAAACATTCATGTCACTGAAAGAGAAATGGAAATGAGAACATTCGA 3060  
Db 3120 GAAAACTTTGAGAAACATTCATGTCACTGAAAGAGAAATGGAAATGAGAACATTCGA 3179  
OY 3061 AGTACAGTGAAGCAATTAAGCCGTATATAACATTAAGAAATGTTTTTAAAGAGCCAGC 3120  
Db 3180 AGTACAGTGAAGCAATTAAGCCGTATATAACATTAAGAAATGTTTTTAAAGAGCCAGC 3239  
OY 3121 TCAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3180  
Db 3240 TCAAGCAATATTAATGAAGTAGTTCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3299

OY 3181 ATAGTTCCAGTATGAAGAAACATTCAGCAGAACTAGTGAAGACAGAGGCCCCAAATTTG 3240  
Db 3300 ATAGTTCCAGTATGAAGAAACATTCAGCAGAACTAGTGAAGACAGAGGCCCCAAATTTG 3359  
OY 3241 AATGCTATCTTATAGATTAGGGGTTTTGCAACCTGAGGCTATATAACAAAGTCTTCGGA 3300  
Db 3360 AATGCTATCTTATAGATTAGGGGTTTTGCAACCTGAGGCTATATAACAAAGTCTTCGGA 3419  
OY 3301 ACTAATTTAGCATCCCTGAATTAAGAAAGCAAGATATGAAGAGTGTTCAGACTGT 3360  
Db 3420 ACTAATTTAGCATCCCTGAATTAAGAAAGCAAGATATGAAGAGTGTTCAGACTGT 3479  
OY 3361 AATCAGATTTCTCCCATATCTGATTTAGATTAAGAAACCCATGGAAGT 3420  
Db 3480 AATCAGATTTCTCCCATATCTGATTTAGATTAAGAAACCCATGGAAGT 3539  
OY 3421 CATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGTGAATAAG 3480  
Db 3540 CATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGTGAATAAG 3599  
OY 3481 GAAAGTACTAGTTTGTCTGAAATGACATTAAGAAAGTGTGCTGTTTTAGCAAAAGC 3540  
Db 3600 GAAAGTACTAGTTTGTCTGAAATGACATTAAGAAAGTGTGCTGTTTTAGCAAAAGC 3659  
OY 3541 GTCCAGAAAGAGAGCTTAGCAGAGTCCATCCCTTCCACCATACATTTGGCTCAG 3600  
Db 3660 GTCCAGAAAGAGAGCTTAGCAGAGTCCATCCCTTCCACCATACATTTGGCTCAG 3719  
OY 3601 GGTTCACGAAGAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTATCTAGTAGAT 3660  
Db 3720 GGTTCACGAAGAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTATCTAGTAGAT 3779  
OY 3661 GAAGACCTCCCTGCTCCCAACACTTGTATTGTGAAGTAACAAATACCTTCAG 3720  
Db 3780 GAAGACCTCCCTGCTCCCAACACTTGTATTGTGAAGTAACAAATACCTTCAG 3839  
OY 3721 TCTACTAGGCATAGCACCGCTTGCTACGAGTCTGTCTAAGAACACAGAGAGAAATTA 3780  
Db 3840 TCTACTAGGCATAGCACCGCTTGCTACGAGTCTGTCTAAGAACACAGAGAGAAATTA 3899  
OY 3781 TTATCATTTGAACAATAGCTTAATATGACTGCACTAACCGAGTAATTTGGCAAGGCTAT 3840  
Db 3900 TTATCATTTGAACAATAGCTTAATATGACTGCACTAACCGAGTAATTTGGCAAGGCTAT 3959  
OY 3841 CAGGAACATCACCTTAGTGAAGAAACAAAGTTGCTGACTGTTTCTTCACAGTGC 3900  
Db 3960 CAGGAACATCACCTTAGTGAAGAAACAAAGTTGCTGACTGTTTCTTCACAGTGC 4019  
OY 3901 AGTGAATTTGGAAGACTTGACTGCAAAATACAAACACCCAGATCCCTTCTGATTTGTTCT 3960  
Db 4020 AGTGAATTTGGAAGACTTGACTGCAAAATACAAACACCCAGATCCCTTCTGATTTGTTCT 4079  
OY 3961 TCCAAACAAATGAGCATGACTGTAAGAGCCAGGAGTGTGCTGAGTGAACAAGATTTG 4020  
Db 4080 TCCAAACAAATGAGCATGACTGTAAGAGCCAGGAGTGTGCTGAGTGAACAAGATTTG 4139  
OY 4021 GTTTCAGATGATGAAGAAAGAGAGGAGGCTTGGAGAGAAATATCAAGAAAGACCAAGC 4080  
Db 4140 GTTTCAGATGATGAAGAAAGAGAGGAGGCTTGGAGAGAAATATCAAGAAAGACCAAGC 4199  
OY 4081 ATGATTTCAAACCTTAGAGTAAGCAGCATCTGGGTGTGAGTGAAGACAGGCTCTCGAA 4140  
Db 4200 ATGATTTCAAACCTTAGAGTAAGCAGCATCTGGGTGTGAGTGAAGACAGGCTCTCGAA 4259  
OY 4141 GACTGCTAGAGGCTATCTCTCAGAGTGAACATTTTAACCACTGCGAGAGGATACCAGT 4200  
Db 4260 GACTGCTAGAGGCTATCTCTCAGAGTGAACATTTTAACCACTGCGAGAGGATACCAGT 4319  
OY 4201 CAACATTAACCTGATTAAGCTCAGAGAAATGCTGACATGAAGAGTGTGTTAAGACAG 4260  
Db 4320 CAACATTAACCTGATTAAGCTCAGAGAAATGCTGACATGAAGAGTGTGTTAAGACAG 4379

OY	4261	ATATGGAGCCAGCCTTCTAACAGCTACCCCTCCATCTAAAGTGCCTCTGGCCCTTGAG	4320
Db	4380	CAATGGAGCCAGCCTTCTAACAGCTACCCCTCCATCTAAAGTGCCTCTGGCCCTTGAG	4439
OY	4321	GACCTGGAAATCCAGAAACAAGCAGATCTCAGAAAACAGTATTAACCTTCACAGAAAGT	4380
Db	4440	GACCTGGAAATCCAGAAACAAGCAGATCTCAGAAAACAGTATTAACCTTCACAGAAAGT	4499
OY	4381	AGTGAATACCTTAAAGCCAGAAATCCAGAAAGCCTTTCTGCTGACAAGTTTGAAGTGTCT	4440
Db	4500	AGTGAATACCTTAAAGCCAGAAATCCAGAAAGCCTTTCTGCTGACAAGTTTGAAGTGTCT	4559
OY	4441	GCAATAGTCTTACCAGTAAAAAATAAAGAACACAGAGTGGAAAGTATCCCTCTTAAA	4500
Db	4560	GCAATAGTCTTACCAGTAAAAAATAAAGAACACAGAGTGGAAAGTATCCCTCTTAAA	4619
OY	4501	TGCCATCATATAGATAGTGGTGTACATGCACAGTTGCTCTGGAGCTCTCAAAATAGA	4560
Db	4620	TGCCATCATATAGATAGTGGTGTACATGCACAGTTGCTCTGGAGCTCTCAAAATAGA	4679
OY	4561	AACTAACCCATCTCAAGAGAGAGCTATTAAAGTTGTGATGTGAGAGCAACAGCTGGAA	4620
Db	4680	AACTAACCCATCTCAAGAGAGAGCTATTAAAGTTGTGATGTGAGAGCAACAGCTGGAA	4739
OY	4621	GAGCTGTGGCCACACAGATTTTACGCGAAACATCTTACTTGGCCAAAGATCTTGAAGGA	4680
Db	4740	GAGCTGTGGCCACACAGATTTTACGCGAAACATCTTACTTGGCCAAAGATCTTGAAGGA	4799
OY	4681	ACCCCTTAACCTGGAATCTGGAATAGCCTCTTCTATGACCCCTGATCTGATCTCTCT	4740
Db	4800	ACCCCTTAACCTGGAATCTGGAATAGCCTCTTCTATGACCCCTGATCTGATCTCTCTCT	4859
OY	4741	GAAAGACAGAGCCCGAGAGTCACTGCTGTGGCAACATACCATTCTCAACCTCGCATTG	4800
Db	4860	GAAAGACAGAGCCCGAGAGTCACTGCTGTGGCAACATACCATTCTCAACCTCGCATTG	4919
OY	4801	AAAGTCCCCCAATGGAAGTTGAGAGATCTGCCCCAGAGTCCAGTGCCTCATTACTACT	4860
Db	4920	AAAGTCCCCCAATGGAAGTTGAGAGATCTGCCCCAGAGTCCAGTGCCTCATTACTACT	4979
OY	4861	GATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGAGGAGAGCCAGATTTACA	4920
Db	4980	GATACTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGAGGAGAGCCAGATTTACA	5039
OY	4921	GCTTCAACAGAAAGGTCACACAAAAGATGTCATGGTGTGTCTGGCCTGACCCCGAAA	4980
Db	5040	GCTTCAACAGAAAGGTCACACAAAAGATGTCATGGTGTGTCTGGCCTGACCCCGAAA	5099
OY	4981	GAATTTATGCTCGCTGTACAAAGTTTGGCCAGAAACACACATCACTTTAACTAATCTAAT	5040
Db	5100	GAATTTATGCTCGCTGTACAAAGTTTGGCCAGAAACACACATCACTTTAACTAATCTAAT	5159
OY	5041	ACTGAAGAGCATCTCATGTTGTTATGAAAAACATCTGTGTTGTGTGAAGCGACA	5100
Db	5160	ACTGAAGAGCATCTCATGTTGTTATGAAAAACATCTGTGTTGTGTGTGAAGCGACA	5219
OY	5101	CTGAATAATTTTTCAGAAATTCGCGGAGAGAAATGGGTAGTTAGCTATTTCTGGGTACCC	5160
Db	5220	CTGAATAATTTTTCAGAAATTCGCGGAGAGAAATGGGTAGTTAGCTATTTCTGGGTACCC	5279
OY	5161	CAGCTCTATTAAAGAAAGAAAAATCTCAATAGCATATTTTGAAGTCAGAGAGATGTG	5220
Db	5280	CAGCTCTATTAAAGAAAGAAAAATCTCAATAGCATATTTTGAAGTCAGAGAGATGTG	5339
OY	5221	GTCATATGGAAGAAACCCACCAAGGTCCAAAGGACACAGATATCCCGAGACAGAAATGC	5280
Db	5340	GTCATATGGAAGAAACCCACCAAGGTCCAAAGGACACAGATATCCCGAGACAGAAATGC	5399
OY	5281	TTCAAGGGGGCTAGAAATCTGTGTATGGGCCCTTCCACCAATGCCACAGATCAACTG	5340
Db	5400	TTCAAGGGGGCTAGAAATCTGTGTATGGGCCCTTCCACCAATGCCACAGATCAACTG	5459
OY	5341	GAATGATGGTACAGCTGTGTGTGTCTGTGTGTGAAGGACCTTTCAATCATCAACCTTT	5400

Db	5460	GAATGGATGGTACACCTTGTGGTCTTCTGTGGTGAAGAGACTTTCATCATTACACCTT	5519
QY	5401	GGCAGAGTGTCCACCCCAATTGTGGTTGTGCAGCCAGATGCGTGGACAGAGACAAATGGC	5460
Db	5520	GGCAGAGTGTCCACCCCAATTGTGGTTGTGCAGCCAGATGCGTGGACAGAGACAAATGGC	5579
QY	5461	TTCCATGCAATTGGGCGAGATGTGTGAGGCACCTGTGTGTGACCCAGACAGATGGGTGTGGAC	5520
Db	5580	TTCCATGCAATTGGGCGAGATGTGTGAGGCACCTGTGTGTGACCCAGACAGATGGGTGTGGAC	5539
QY	5521	AGTGTAGACCTCTTACCAATGGCCAGAGGTGGACACCTTACCTGATACCCCAATGCCCCAC	5580
Db	5640	AGTGTAGACCTCTTACCAATGGCCAGAGGTGGACACCTTACCTGATACCCCAATGCCCCAC	5699
QY	5581	AGCCACTTAC 5589	
Db	5700	AGCCACTTAC 5708	
RESULT 10			
LOCUS	AR008159	5914 bp	DNA
DEFINITION	Sequence 1 from patent US 5753441.		Linear
ACCESSION	AR008159		
VERSION	AR008159.1	GI:3967268	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5914)		
AUTHORS	Skolnick, M.H., Goldgar, D.E., Miki, Y., Swenson, J., Kamb, A.,		
	Hardman, K.D., Shattuck-Eidens, D.M., Tavtigian, S.V., Wiseman, R.W.		
	and Futreal, P. Andrew.		
TITLE	170-linked breast and ovarian cancer susceptibility gene		
JOURNAL	Patent: US 5753441-A 1 19-MAY-1998;		
FEATURES	Location/Qualifiers		
	1..5914		
	Source		
	1..5914		
	/organism="unknown"		
BASE COUNT	2006 a 1156 c 1316 g 1436 t		
ORIGIN			
Query Match	99.98;	Score 5585.8;	DB 6; length 5914;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 5587;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	1	ATGGAATTATCTGCTCTTGGGTGAGAGAACTACAAATGTCATTATGTCATGACGAAA	60
Db	120	ATGGAATTATCTGCTCTTGGGTGAGAGAACTACAAATGTCATTATGTCATGACGAAA	179
QY	61	ATCTTAGAGTGTCCCATCTGTCTGGAGTGTGATCAAGAACCTGTCTCCAAAGTGTGAC	120
Db	180	ATCTTAGAGTGTCCCATCTGTCTGGAGTGTGATCAAGAACCTGTCTCCAAAGTGTGAC	239
QY	121	CACATATTTTGCATAATTTTGCATCTGAACCTTCACACGAGAAAGGGCTTCACAG	180
Db	240	CACATATTTTGCATAATTTTGCATCTGAACCTTCACACGAGAAAGGGCTTCACAG	299
QY	181	TGTCCTTTAGTAGAATGATATTAACCAAAAGAGCCTTACAAAGAAAGTATGAGATTAGT	240
Db	300	TGTCCTTTAGTAGAATGATATTAACCAAAAGAGCCTTACAAAGAAAGTATGAGATTAGT	359
QY	241	CAACTTGTGGAAGAGCTATTTGCAAAATCATTTGTCTTTGACCTTGCACACAGTTTGGAG	300
Db	360	CAACTTGTGGAAGAGCTATTTGCAAAATCATTTGTCTTTGACCTTGCACACAGTTTGGAG	419
QY	301	TATGCAAAACAGCTATTAATTTTGCAAAAAGAAATTAATCTCCTGAACATCTAAAGAT	360
Db	420	TATGCAAAACAGCTATTAATTTTGCAAAAAGAAATTAATCTCCTGAACATCTAAAGAT	479
QY	361	GAATTTTCTATCATCCAAAGATATGGGTACAGAAACCGTGCACAAAGACTTTCACAGT	420
Db	480	GAATTTTCTATCATCCAAAGATATGGGTACAGAAACCGTGCACAAAGACTTTCACAGT	539

421 GAACCCGAAATCTCTTCTTGAGAAACGCTCAGTGTCCAACTCTCTAACCTTGA 480  
|||||  
540 GAACCCGAAATCTCTTCTTGAGAAACGCTCAGTGTCCAACTCTCTAACCTTGA 539  
481 ACTGTGAGACTCTGAGCAACAGCAGCGATACAACTCTAAAGAGCTGTCTACAT 540  
600 ACTGTGAGACTCTGAGCAACAGCAGCGATACAACTCTAAAGAGCTGTCTACAT 659  
541 GAATTGGGCTGATCTCTGAGATACCGTTAATAGGCACTATTGCACTGTGGGA 600  
|||||  
660 GAATTGGGCTGATCTCTGAGATACCGTTAATAGGCACTATTGCACTGTGGGA 719  
601 GATCAAGATTTGTACAAATCAACCCCTCAAGGAAACGAGGATGAATTCAGTTGATTC 650  
720 GATCAAGATTTGTACAAATCAACCCCTCAAGGAAACGAGGATGAATTCAGTTGATTC 779  
661 GCAAAAAGGCTGCTGTGTAATTTTCTGAGCGGATGTAACAAATCTGACATCA 720  
780 GCAAAAAGGCTGCTGTGTAATTTTCTGAGCGGATGTAACAAATCTGACATCA 839  
721 CCCAGTAATATGATTTGAACACCACTGGAAGCGTGAGCGCATCCAGAAAG 780  
840 CCCAGTAATATGATTTGAACACCACTGGAAGCGTGAGCGCATCCAGAAAG 899  
781 TATCAGGGTAGTCTGTTTCAAACTTGATGTGAGCCATGTGGCAAAATCTCATGCC 840  
900 TATCAGGGTAGTCTGTTTCAAACTTGATGTGAGCCATGTGGCAAAATCTCATGCC 959  
841 AGCTATTACAGCATGAGCAACAGCAGTTTATTACTACTAAGACAGATGAATGTGA 900  
960 AGCTATTACAGCATGAGCAACAGCAGTTTATTACTACTAAGACAGATGAATGTGA 1019  
901 AAGGCTGAATTTGTATATAAAGCAACAGCCTGGCTTGCAAGAGGCCAATACAGA 960  
1020 AAGGCTGAATTTGTATATAAAGCAACAGCCTGGCTTGCAAGAGGCCAATACAGA 1079  
961 TGGGCTGGAAGTAAGSAAACATGTATGATAGGCGGACTCCAGCAGAAAAAAGTA 1020  
1080 TGGGCTGGAAGTAAGSAAACATGTATGATAGGCGGACTCCAGCAGAAAAAAGTA 1139  
1021 GATCTGAATGCTGATCCCTGTGTGAGAAAAAAGATGCAATAGCAGAACTGCCATGC 1080  
1140 GATCTGAATGCTGATCCCTGTGTGAGAAAAAAGATGCAATAGCAGAACTGCCATGC 1199  
1081 TCAGAGAACTCTAGAGATCTAGAGATGTTCTTGATTAACACTAATAGCAGATTGAG 1140  
1200 TCAGAGAACTCTAGAGATCTAGAGATGTTCTTGATTAACACTAATAGCAGATTGAG 1259  
1141 AAAGTTAATGAGTGTCTTCCAGAGTGATGAACCTGTAGTTCTGATGACTCATGAT 1200  
1260 AAAGTTAATGAGTGTCTTCCAGAGTGATGAACCTGTAGTTCTGATGACTCATGAT 1319  
1201 GGGGAGTCTGATCAAAATGCCAAAGTGTGATTTGAGAGTCTTAAATAGTATGAT 1260  
1320 GGGGAGTCTGATCAAAATGCCAAAGTGTGATTTGAGAGTCTTAAATAGTATGAT 1379  
1261 GAATTTCTGCTTCTCAGAGAAATAGACTACTGGCCAGTATCTCATGGGCTTTA 1320  
1380 GAATTTCTGCTTCTCAGAGAAATAGACTACTGGCCAGTATCTCATGGGCTTTA 1439  
1321 ATATGTAAAGTGAAGAGTCTCAATCAAGTATGAGAGATTAATTTAGACAAATA 1380  
1440 ATATGTAAAGTGAAGAGTCTCAATCAAGTATGAGAGATTAATTTAGACAAATA 1499  
1381 TTTGGGAAAACCTATCGAGAGAGAGCAGCCCTCCCACTTAAGCCATTAAGTAAAT 1440  
1500 TTTGGGAAAACCTATCGAGAGAGAGCAGCCCTCCCACTTAAGCCATTAAGTAAAT 1559  
1441 CTAATTATAGAGCACTTTGTCTAGGACAGATTAATCAAGAGGCTCCCTCACAAT 1500  
1560 CTAATTATAGAGCACTTTGTCTAGGACAGATTAATCAAGAGGCTCCCTCACAAT 1619

1501 AATTAAGGCTAAAGAGAGACCTACATCAGGCTTCACTCTGAGGATTTTATCAGAAA 1560  
|||||  
1620 AATTAAGGCTAAAGAGAGACCTACATCAGGCTTCACTCTGAGGATTTTATCAGAAA 1679  
1561 GCAGATTTGGCACTTCAAAAGACTCCTGAAATGATTAATCAGGAACTTACCAAGAG 1620  
1680 GCAGATTTGGCACTTCAAAAGACTCCTGAAATGATTAATCAGGAACTTACCAAGAG 1739  
1621 CAGATGTCTAGTATGATTAATTAATAGGCTGATGAGATTAATCAAAAGAGTAT 1680  
1740 CAGATGTCTAGTATGATTAATTAATAGGCTGATGAGATTAATCAAAAGAGTAT 1799  
1681 TCTATTAGATGAGAAAAATCCTAACCCATAGATCACTGAAAAAGATCTCTTC 1740  
1800 TCTATTAGATGAGAAAAATCCTAACCCATAGATCACTGAAAAAGATCTCTTC 1859  
1741 AAACGAAAGCTGACCTTATAGCAGCAGTATAGCAATATGGAATCTGAATTAATATC 1800  
1860 AAACGAAAGCTGACCTTATAGCAGCAGTATAGCAATATGGAATCTGAATTAATATC 1919  
1801 CACATTTCAAAAGCAGCTTAAAAAGATAGGCTGAGAGGAGTCTTACCGAGCATAT 1860  
1920 CACATTTCAAAAGCAGCTTAAAAAGATAGGCTGAGAGGAGTCTTACCGAGCATAT 1979  
1861 CATGCGCTTGAACCTAGTACGTAGTAATCTAAGCCCACTAATTTGATCTGAATTCGA 1920  
1980 CATGCGCTTGAACCTAGTACGTAGTAATCTAAGCCCACTAATTTGATCTGAATTCGA 2039  
1921 ATTGATAGTGTCTTACGAGTATAGAGATTAAGAAAAAATATACACCAATTCGCAATC 1980  
2040 ATTGATAGTGTCTTACGAGTATAGAGATTAAGAAAAAATATACACCAATTCGCAATC 2099  
1981 AGGCAAGAGAAACCTACACATCATGAGGTAAGAAACCTGCAACTGAGAGCCAAAG 2040  
2100 AGGCAAGAGAAACCTACACATCATGAGGTAAGAAACCTGCAACTGAGAGCCAAAG 2159  
2041 AGTAACAGCCAAATGAGACAGCAAGTAAGAAACATGACAGTATCTTCCAGAGCTG 2100  
2160 AGTAACAGCCAAATGAGACAGCAAGTAAGAAACATGACAGTATCTTCCAGAGCTG 2219  
2101 AAGTTAAACAAATGACCTGTTCTTTTACTAGTGTCAATACAGAGTAATTAAGAA 2160  
2220 AAGTTAAACAAATGACCTGTTCTTTTACTAGTGTCAATACAGAGTAATTAAGAA 2279  
2161 TTTGCAATCTAGCCTTCCAGAGAGAAAAAGAAAGTAAGTAAGTAAGTAAGTAAG 2220  
2280 TTTGCAATCTAGCCTTCCAGAGAGAAAAAGAAAGTAAGTAAGTAAGTAAGTAAG 2339  
2221 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGCTTTGCCAACT 2280  
2340 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAAGAGGCTTTGCCAACT 2399  
2281 GAAAGATCTGTAGAGTATGAGTATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTG 2340  
2400 GAAAGATCTGTAGAGTATGAGTATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTG 2459  
2341 GAAAGATCTGTAGAGTATGAGTATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTG 2400  
2460 GAAAGATCTGTAGAGTATGAGTATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTG 2519  
2401 TGTGTAGTCACTGTGACACATTTGAAAAACCCCAAGGACATAATTCATGTTGTTCCAA 2460  
2520 TGTGTAGTCACTGTGACACATTTGAAAAACCCCAAGGACATAATTCATGTTGTTCCAA 2579  
2461 GATTAATGAAATGACACAGAGAGCTTTTAAGTATCATTTGGGACATGAAGTTAACACAGT 2520  
2580 GATTAATGAAATGACACAGAGAGCTTTTAAGTATCATTTGGGACATGAAGTTAACACAGT 2639  
2521 CGGGAACAGCATGGAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2580  
2640 CGGGAACAGCATGGAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2639  
2581 TTCAAGCTTTCAAGGCGCCAGTCAATTTGCTGTTTCAATTCAGGAATTCAGAGAGAG 2640

```
|||||
Db 2700 TTCAAGTTTCAAGCGCCAGTCATTTGCTCCGTTTCAATCAGAAATGAGAG 2759
QY 2641 GATNGGCAACATTTCTGCCCACCTCGGGTCTTAAAGAAACAAAGTCACT 2700
Db 2760 GAATGTGCAACATTTCTGCCCACCTCGGGTCTTAAAGAAACAAAGTCACT 2819
QY 2701 TTGAATGTGCAACAAAGAAAGAAATCAAGAAAGAAATGATGATCAAGCCCTGTA 2760
Db 2820 TTGATATGTACAAAGAAAGAAATCAAGAAAGAAATGATGATCAAGCCCTGTA 2879
QY 2761 CAGACATTAATATCACTGAGGCTTCTGTTGGTGTGAGAAAGTAAGCCAGTTGAT 2820
Db 2880 CAGACATTAATATCACTGAGGCTTCTGTTGGTGTGAGAAAGTAAGCCAGTTGAT 2939
QY 2821 AATGCCAAATGTAGTATCAAGAAAGGCTAGGTTTGTCTATCATCTCAGTTCAGAGC 2880
Db 2940 AATGCCAAATGTAGTATCAAGAAAGGCTAGGTTTGTCTATCATCTCAGTTCAGAGC 2999
QY 2881 AAGGAACTGCACTATTACTCCAAATTAACATGAGACTTTTACAAACCCATATCGATA 2940
Db 3000 AAGGAACTGCACTATTACTCCAAATTAACATGAGACTTTTACAAACCCATATCGATA 3059
QY 2941 CCACCACTTTTCCCATCAAGTCATTTGTTAAATTAATGTAAAGAAATCTGCTAGAG 3000
Db 3060 CCACCACTTTTCCCATCAAGTCATTTGTTAAATTAATGTAAAGAAATCTGCTAGAG 3119
QY 3001 GAAATCTTTAGAGACATTTAATGTCCCTGAAAGAGAAATGGGAAATGAGAACTTCCA 3060
Db 3120 GAAATCTTTAGAGACATTTAATGTCCCTGAAAGAGAAATGGGAAATGAGAACTTCCA 3179
QY 3061 AGTACAGTGAACAATTTAGCCCTTAATTAATGAGAAATGTTTAAAGAGCCAGC 3120
Db 3180 AGTACAGTGAACAATTTAGCCCTTAATTAATGAGAAATGTTTAAAGAGCCAGC 3239
QY 3121 TCAGCAATATTAATGAGTAGGTTCCAGTACTAATGAGTAGGAGTCCAGTATTAATGAA 3180
Db 3240 TCAGCAATATTAATGAGTAGGTTCCAGTACTAATGAGTAGGAGTCCAGTATTAATGAA 3299
QY 3181 ATAGGTTCCAGTATGAAACATTTCAAGCAAGTACTAGTAAACAGAGGCCCAAAATG 3240
Db 3300 ATAGGTTCCAGTATGAAACATTTCAAGCAAGTACTAGTAAACAGAGGCCCAAAATG 3359
QY 3241 AATGCTATGCTAGTATGAGGTTTTCGCAACCTGAGGCTATTAACAAATCTTCTCGA 3300
Db 3360 AATGCTATGCTAGTATGAGGTTTTCGCAACCTGAGGCTATTAACAAATCTTCTCGA 3419
QY 3301 AGTAAATGTAGCAATCTGAAATPAAAAAGCAAGATATGAGAGTAGTTCAAGCTGT 3360
Db 3420 AGTAAATGTAGCAATCTGAAATPAAAAAGCAAGATATGAGAGTAGTTCAAGCTGT 3479
QY 3361 AATACGATTTCTCTCATATCTGATTTTCAGATTAATTAAGACGCTATGGAAGTAGT 3420
Db 3480 AATACGATTTCTCTCATATCTGATTTTCAGATTAATTAAGACGCTATGGAAGTAGT 3539
QY 3421 CATGCAATCTAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAAATAG 3480
Db 3540 CATGCAATCTAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGTTGAAATAG 3599
QY 3481 GAAGATATAGTATTTGCTGAAATGACATTAAGGAAATGTTCTGCTTTTACCAAAAGC 3540
Db 3600 GAAGATATAGTATTTGCTGAAATGACATTAAGGAAATGTTCTGCTTTTACCAAAAGC 3659
QY 3541 GTCCAGAAAGAGAGCTTAGCAGAGTCCCTAGGCCCTTTACCCATACACATTTGGCTCAG 3600
Db 3660 GTCCAGAAAGAGAGCTTAGCAGAGTCCCTAGGCCCTTTACCCATACACATTTGGCTCAG 3719
QY 3601 GGTATCCGAGAGAGGCGCAGAAATTTAGATCTCCAGAGAGAACTATATAGAGAGAT 3660
Db 3720 GGTATCCGAGAGAGGCGCAGAAATTTAGATCTCCCTCAGAGAGAACTATATAGAGAGAT 3779
QY 3661 GAAGAGCTTCCGCTTCCAAACCTGTTATTTGTTAAAGTAAACATATACCTTCTCAG 3720
|||||

Db 3780 GAAGAGCTTCCGCTTCCAAACCTGTTATTTGTTAAAGTAAACATATACCTTCTCAG 3839
QY 3721 TCTACTAGGATATGACACCGTTGCTACGAGAGTGTGCTAAGAACAGAGAGAAATTTA 3780
Db 3840 TCTACTAGGATATGACACCGTTGCTACGAGAGTGTGCTAAGAACAGAGAGAAATTTA 3899
QY 3781 TTATCATTTGAAGATATGATTTAAATGATGAGTAAACAGGTAATATTTGGCAAGGATCT 3840
Db 3900 TTATCATTTGAAGATATGATTTAAATGATGAGTAAACAGGTAATATTTGGCAAGGATCT 3959
QY 3841 CAGAACATCACTTTAGTGAAGAAACAAATGTTTCTGATAGCTTTTCTTACAGTGC 3900
Db 3960 CAGAACATCACTTTAGTGAAGAAACAAATGTTTCTGATAGCTTTTCTTACAGTGC 4019
QY 3901 AGTGAATTTGAAGACTGACTGCAGAAATACAAACCCAGGATCTTCTTGTATGTTCT 3960
Db 4020 AGTGAATTTGAAGACTGACTGCAGAAATACAAACCCAGGATCTTCTTGTATGTTCT 4079
QY 3961 TCCAAACAAATGAGGCAATCAGTCTGAAGCCAGGAGTGTGCTGAGTGAACAAGATTC 4020
Db 4080 TCCAAACAAATGAGGCAATCAGTCTGAAGCCAGGAGTGTGCTGAGTGAACAAGATTC 4139
QY 4021 GTTTCAGATGATGAAGAGAGAGGAGCGGCTTGAGAGAAATATCAAGAGAGCAAGC 4080
Db 4140 GTTTCAGATGATGAAGAGAGAGGAGCGGCTTGAGAGAAATATCAAGAGAGCAAGC 4199
QY 4081 ATGATTTCAACTTATAGTGAAGCAGCATCTGGGTGTAGAGTGAACAAACCTCTCGAA 4140
Db 4200 ATGATTTCAACTTATAGTGAAGCAGCATCTGGGTGTAGAGTGAACAAACCTCTCGAA 4259
QY 4141 GACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGATACCATG 4200
Db 4260 GACTGCTCAGGCGTATCTCTCAGAGTGAATTTTAACTCAGCAGAGAGATACCATG 4319
QY 4201 CAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAGTGAAGCTGTGTTAGAACAG 4260
Db 4320 CAACATTAACCTGATTAAGCTCCAGCAGAGAAATGGCTGAGTGAAGCTGTGTTAGAACAG 4379
QY 4261 CATGGAGCGAGCTTCTAAGTACCTACCTTCCATCTAATGATGATCTTCTGCCCTGAG 4320
Db 4380 CATGGAGCGAGCTTCTAAGTACCTACCTTCCATCTAATGATGATCTTCTGCCCTGAG 4439
QY 4321 GACCTCGGAATCCAGAACACACATCAGAAACAAAGCACTATTAATCTCAGAGAAAGT 4380
Db 4440 GACCTCGGAATCCAGAACACACATCAGAAACAAAGCACTATTAATCTCAGAGAAAGT 4499
QY 4381 AGTGAATACCTTATTAAGCCAGAAATCCAGAGGCCCTTCTGCTGACAAAGTTGAGGTGCT 4440
Db 4500 AGTGAATACCTTATTAAGCCAGAAATCCAGAGGCCCTTCTGCTGACAAAGTTGAGGTGCT 4559
QY 4441 GCAGATAGTCTTACCAATTAAGAAATTAAGAACCCAGAGTGAAGAGTATCCCTCTTAAA 4500
Db 4560 GCAGATAGTCTTACCAATTAAGAAATTAAGAACCCAGAGTGAAGAGTATCCCTCTTAAA 4619
QY 4501 TGCCCATCATTTAATGATGATGTTGTTGATCATGACAGTGTCTGGAGATCTTTCAGAAATAGA 4560
Db 4620 TGCCCATCATTTAATGATGATGTTGTTGATCATGACAGTGTCTGGAGATCTTTCAGAAATAGA 4679
QY 4561 AACTTACCATCTCAAGAGAGCTCATTAAAGTTGTTGATGTGAGAGAGCAACAGCTGGAA 4620
Db 4680 AACTTACCATCTCAAGAGAGCTCATTAAAGTTGTTGATGTGAGAGAGCAACAGCTGGAA 4739
QY 4621 GAGTCTGGGCGCACAGATTTGAGCGGAAACATTTACTTGCCAAAGGCAAGATCTAGAGGGA 4680
Db 4740 GAGTCTGGGCGCACAGATTTGAGCGGAAACATTTACTTGCCAAAGGCAAGATCTAGAGGGA 4799
QY 4681 ACCCTTACCTGGAATCTGGAATCAGGCTCTCTCTGATATACCTGATATCTGATCTTCT 4740
Db 4800 ACCCTTACCTGGAATCTGGAATCAGGCTCTCTCTGATATACCTGATATCTGATCTTCT 4859
QY 4741 GAAGACAGAGCCCGAGAGTCTGTTGGCAACATATACATCTTCAACCTCTGCAATG 4800
Db 4860 GAAGACAGAGCCCGAGAGTCTGTTGGCAACATATACATCTTCAACCTCTGCAATG 4919
```

QY	4801	AAAGTTCCTCCCAATGGAATGGTGCAGAAATCTGCTCCAGAGATCCAGCTGCTGCTCATATCACT	4860
Db	4920	AAAGTTCCTCCCAATGGAATGGTGCAGAAATCTGCTCCAGAGATCCAGCTGCTGCTCATATCACT	4979
QY	4861	GATATCTGCTGGGTATATATGCAATGGAAGAGTGTGAGCAGAGGAGAAAGCCAGAAATTGACA	4920
Db	4980	GATACTGCTGGGTATATATGCAATGGAAGAGTGTGAGCAGAGGAGAAAGCCAGAAATTGACA	5039
QY	4921	GCTTCACAGAAAGGGTCAACAAAAGATGTTCATGGTGTGTGCTGGCTTGACCCAGAA	4980
Db	5040	GCTTCACAGAAAGGGTCAACAAAAGATGTTCATGGTGTGTGCTGGCTTGACCCAGAA	5099
QY	4981	GAATTTATGTCGCTGTACAGTTGGCCAGAAAACCCAGATCAGTTTAACCTAATCTAATT	5040
Db	5100	GAATTTATGTCGCTGTGTACAGTTGGCCAGAAAACCCAGATCAGTTTAACCTAATCTAATT	5159
QY	5041	ACTGAAGAGACTACTCATGTGTTATGAAAACAGATGCTGAGTTGTGTGTGAACGGACA	5100
Db	5160	ACTGAAGAGACTACTCATGTGTTATGAAAACAGATGCTGAGTTGTGTGTGAACGGACA	5219
QY	5101	CTGAATATTTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATTTCTGGGTGACC	5160
Db	5220	CTGAATATTTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATTTCTGGGTGACC	5279
QY	5161	CAGCTATATTAAGAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTG	5220
Db	5280	CAGCTATATTAAGAAAGAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTG	5339
QY	5221	GTCAATGGAAGAAAACCAACCAAGGTCCTCAAGGAGCAGAGAGAAATCCACAGACAGAAAGATC	5280
Db	5340	GTCAATGGAAGAAAACCAACCAAGGTCCTCAAGGAGCAGAGAGAAATCCACAGAGAAAGATC	5399
QY	5281	TTTCAGGGGGCTAGAAATCTGTTGCTATGAGGGCCCTTACCAACATATGCCACAGATCAACTG	5340
Db	5400	TTTCAGGGGGCTAGAAATCTGTTGCTATGAGGGCCCTTACCAACATATGCCACAGATCAACTG	5459
QY	5341	GAATGGAATGCTACAGCTGTGTGTGCTTCTGTGTGAAGAGCTTTTCATCTTACACCTT	5400
Db	5460	GAATGGAATGCTACAGCTGTGTGTGCTTCTGTGTGAAGAGCTTTTCATCTTACACCTT	5519
QY	5401	GGCACAGGTGCCACCCAAATTTGTGTTGTGACGCACAGATGCTGTGACAGAGACAAATGGC	5460
Db	5520	GGCACAGGTGCCACCCAAATTTGTGTTGTGACGCACAGATGCTGTGACAGAGACAAATGGC	5579
QY	5461	TTTCATGCAATTGGGGCAGATGTGTGAGGCACCTGTGGGACCCAGACAGATGGGTGTGGAC	5520
Db	5580	TTTCATGCAATTGGGGCAGATGTGTGAGGCACCTGTGGGACCCAGACAGATGGGTGTGGAC	5639
QY	5521	AGTGTAGACACTCTACAGTGTGCAGAGAGCTGACACCTAAGCTGAATACCCACAGATCCCCAC	5580
Db	5640	AGTGTAGACACTCTACAGTGTGCAGAGAGCTGACACCTAAGCTGAATACCCACAGATCCCCAC	5699
QY	5581	AGCCACTTAC 5589	
Db	5700	AGCCACTTAC 5708	

RESULT	11		
LOCUS	AR136942	5914 bp	DNA
DEFINITION	AR136942		linear
ACCESSION	Sequence 1 from patent US 6162897.		PAT 16-JUN-2001
VERSION	AR136942		
KEYWORDS	AR136942.1	GI:14478192	
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 5914)		
	Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,		
	Haashman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.		
	and Futreal,P.Andrew.		
TITLE	17q-linked breast and ovarian cancer susceptibility gene		

JOURNAL	Patent: US 6162897-A 1 19-DEC-2000
FEATURES	Location/Qualifiers
SOURCE	1. .5914
	/Organization="unknown"
BASE COUNT	2006 a 1156 c 1316 g 1436
ORIGIN	

Query Match	99.98;	Score 5585.8;	DB 6;	length 5914;
Best Local Similarity	100.08;	Pred. No. 0;		
Matches 5587; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	ATGATATTTATTCGCTCTTGCGGTGAAGACATCAAAATGCTATTAATGCTATGCGAAA	60
Dp		
Dp	120 ATGATATTTATTCGCTCTTGCGGTGAAGACATCAAAATGCTATTAATGCTATGCGAAA	179
QY	ATCTTAGAGAGTCCCATCTGCTGGAGTTGATCAAGAGACCTGCTCCAAAGTGTGAC	120
Dp		
Dp	180 ATCTTAGAGAGTCCCATCTGCTGGAGTTGATCAAGAGACCTGCTCCAAAGTGTGAC	239
QY	121 CACATATTTTGCAATTTTGCAATGCTGCTGAACCTTCTCAACGAGAAAGGGCTTCACAG	180
Dp		
Dp	240 CACATATTTTGCAAAATTTTGCAATGCTGCTGAACCTTCTCAACGAGAAAGGGCTTCACAG	299
QY	181 TGTCCTTTATGTAAGATGATTAATCAACAAAGAGGCTTACAGAAAGTACGATTTACT	240
Dp		
Dp	300 TGTCCTTTATGTAAGATGATTAATCAACAAAGAGGCTTACAGAAAGTACGATTTACT	359
QY	241 CAACCTGTTGAGAGAGCTATTTGAATAATCATTTGTCTTTCAGCTTGACACAGGTTTGAG	300
Dp		
Dp	360 CAACCTGTTGAGAGAGCTATTTGAATAATCATTTGTCTTTCAGCTTGACACAGGTTTGAG	419
QY	301 TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAATTAATCTCTCTGACATCTTAAAGAT	360
Dp		
Dp	420 TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAATTAATCTCTCTGACATCTTAAAGAT	479
QY	361 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACGCTCCAAAGAGCTTCTACAGAT	420
Dp		
Dp	480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACGCTCCAAAGAGCTTCTACAGAT	539
QY	421 GAACCCGAAATCTTTCCTTGCAAGAAACAGTCTCAGTGTCCAACTCTTAACTTTGGA	480
Dp		
Dp	540 GAACCCGAAATCTTTCCTTGCAAGAAACAGTCTCAGTGTCCAACTCTTAACTTTGGA	599
QY	481 ACGTATGAGAACTGTAGGACAAAGACGAGGATTAACCTCAAAAGAGCTGTCTACAT	540
Dp		
Dp	600 ACGTATGAGAACTGTAGGACAAAGACGAGGATTAACCTCAAAAGAGCTGTCTACAT	659
QY	541 GAATTTGGAGTCGATTTCTTCTGAAATATCCCTTATTAAGGCAACTTATTTGAGTGGGA	600
Dp		
Dp	660 GAATTTGGAGTCGATTTCTTCTGAAATATCCCTTATTAAGGCAACTTATTTGAGTGGGA	719
QY	601 GATCAAGAATTTGTTACAAATCACCCCTTAAGGAACCGAGATGAATCAGTTTGGATTCT	660
Dp		
Dp	720 GATCAAGAATTTGTTACAAATCACCCCTTAAGGAACCGAGATGAATCAGTTTGGATTCT	779
QY	661 GCAAAAAAGCGCTGTGGAAATTTCTGAGACGGATTAACAATCTGACATCA	720
Dp		
Dp	780 GCAAAAAAGCGCTGTGGAAATTTCTGAGACGGATTAACAATCTGACATCA	839
QY	721 CCCAGTAATTAATGATTTTGAACACCACTGAGAAAGCTCAGCTGAGAGGCTCAGAAAAAG	780
Dp		
Dp	840 CCCAGTAATTAATGATTTTGAACACCACTGAGAAAGCTCAGCTGAGAGGCTCAGAAAAAG	899
QY	781 TATCAGAGGTAGTCTTGTTTCAAACTTGATGTGAGAGCATGTGGCAAAATCTATGCC	840
Dp		
Dp	900 TATCAGAGGTAGTCTTGTTTCAAACTTGATGTGAGAGCATGTGGCAAAATCTATGCC	959
QY	841 AGCTATTTACACATGAGAACAGCAGTTTATATCTACTTAAAGACAGATGAATAGAA	900
Dp		
Dp	960 AGCTATTTACACATGAGAACAGCAGTTTATATCTACTTAAAGACAGATGAATAGAA	1019
QY	901 AAGCGTGAATTCGTAAATTAAGCAAAACAGCTTGCTTTAGCAAGAGGCCAACATTAACA	960
Dp		



Db 1020 AAGGCTGAATTCGTGAATTAAGCAACAGCCTGGCTTAGCCAGGACCAACATTAACAGA 1079  
Qy TGGGCTGGAAGTAAGAAACATGTAATGATAGCGGAGCTCCAGCAGACAGAAAAAGGTA 1020  
Db TGGGCTGGAAGTAAGAAACATGTAATGATAGCGGAGCTCCAGCAGACAGAAAAAGGTA 1139  
Qy 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGGAATTAAGCAGAACTCCATGC 1080  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGGAATTAAGCAGAACTCCATGC 1139  
Qy 1081 TCAGAGAACTCTAGAGTACTGAAGATGTTCTTGATTAACACTAAATAGCAGCATTCAG 1140  
Db 1200 TCAGAGAACTCTAGAGTACTGAAGATGTTCTTGATTAACACTAAATAGCAGCATTCAG 1259  
Qy 1141 AAGGTAATGAGTGGTTTCCAGAAAGTGAAGTGAAGTGTAGGTTGATGACTGCATGAT 1200  
Db 1260 AAGGTAATGAGTGGTTTCCAGAAAGTGAAGTGAAGTGTAGGTTGATGACTGCATGAT 1319  
Qy 1201 GGGGAGTCTGAATCAAAATGCCAAAGTAGTGTATTTGACGCTTCAATGAGGTAGAT 1260  
Db 1320 GGGGAGTCTGAATCAAAATGCCAAAGTAGTGTATTTGACGCTTCAATGAGGTAGAT 1379  
Qy 1261 GAATATCTGCTGTTCTTTCAGAGAAATAGACTTACTGGCCAGTATCTCATGAGGCTTTA 1320  
Db 1380 GAATATCTGCTGTTCTTTCAGAGAAATAGACTTACTGGCCAGTATCTCATGAGGCTTTA 1439  
Qy 1321 ATATGTAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1380  
Db 1440 ATATGTAAGTGAAGAGTTCACTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1499  
Qy 1381 TTTGGGAAAACCTATCGGAAGAGCAAGCCTCCCACTTAAGCCATGTAACCTGAAAT 1440  
Db 1500 TTTGGGAAAACCTATCGGAAGAGCAAGCCTCCCACTTAAGCCATGTAACCTGAAAT 1559  
Qy 1441 CTATATTTAGAGATTTGTTTACTGAGCAACAGTATTAACAGAGCCTCCCTCACAAT 1500  
Db 1560 CTATATTTAGAGATTTGTTTACTGAGCAACAGTATTAACAGAGCCTCCCTCACAAT 1619  
Qy 1501 AAATTAAGCGTAAAGAGACCTACATCAGGCTTCATCCCTGAGATTTTATCAAGAA 1560  
Db 1620 AAATTAAGCGTAAAGAGACCTACATCAGGCTTCATCCCTGAGATTTTATCAAGAA 1679  
Qy 1561 GCAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAGGAACTAAACCAACGAG 1620  
Db 1680 GCAGATTTGGCAGTTCAAAAAGACTCTGAAATGATTAATCAGGAACTAAACCAACGAG 1739  
Qy 1621 CAGAGTGTCAAGTAGATTAATTAATAGTGTATGATGATTAACAGAAAGGATGAT 1680  
Db 1740 CAGAGTGTCAAGTAGATTAATTAATAGTGTATGATGATTAACAGAAAGGATGAT 1799  
Qy 1681 TCTATTCAGATGAGAAATTCCTAACCCATAGAAATCAGTCCGAAAAAGATGCTTTC 1740  
Db 1800 TCTATTCAGATGAGAAATTCCTAACCCATAGAAATCAGTCCGAAAAAGATGCTTTC 1859  
Qy 1741 AAAACGAAAGCTGAACTATTAAGCAGCATTAAGCAATATGAACTCGAATTAATATC 1800  
Db 1860 AAAACGAAAGCTGAACTATTAAGCAGCATTAAGCAATATGAACTCGAATTAATATC 1919  
Qy 1801 CACAATTCAAAGACCTTAATAAAGAAATAGGCTGAGAGGAAAGTCTTACCAGGCAAT 1860  
Db 1920 CACAATTCAAAGACCTTAATAAAGAAATAGGCTGAGAGGAAAGTCTTACCAGGCAAT 1979  
Qy 1861 CAGCGCTGTAAGTACTAGTAGTAAGTAAAGCCACCTAATTTACTGAAATTCGAA 1920  
Db 1980 CAGCGCTGTAAGTACTAGTAGTAAGTAAAGCCACCTAATTTACTGAAATTCGAA 2039  
Qy 1921 ATTGAATGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAGTAAACAACAATGCCAGTC 1980  
Db 2040 ATTGAATGTTGTTCTAGCAGTGAAGAGATTAAGAAAAAGTAAACAACAATGCCAGTC 2099  
Qy 1981 AGGACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACTGCACCTGAGCCAAAG 2040  
Db 2100 AGGACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACTGCACCTGAGCCAAAG 2159

Qy 2041 AGTAACAAGCCAAATGAACAGACAGTAAGTAAGACATGACATGATCTTCCAGAGCT 2100  
Db 2160 AGTAACAAGCCAAATGAACAGACAGTAAGTAAGACATGACATGATCTTCCAGAGCT 2219  
Qy 2101 AAGTTAACAAATGACACCTGGTTCTTTACTAGTGTTCAAATTAACAGTGAACCTTAAGAA 2160  
Db 2220 AAGTTAACAAATGACACCTGGTTCTTTACTAGTGTTCAAATTAACAGTGAACCTTAAGAA 2279  
Qy 2161 TTTGTCAATCCTACCTTTCCAGAGAAAGAAAGAAAGAAACCTAGAAACGTTAAAGTG 2220  
Db 2280 TTTGTCAATCCTACCTTTCCAGAGAAAGAAAGAAAGAAACCTAGAAACGTTAAAGTG 2339  
Qy 2221 TCTATATGCTGTAAGACCCCAAGAAATCTCATGTTAAGTGAAGAGGTTTCCAACT 2280  
Db 2340 TCTATATGCTGTAAGACCCCAAGAAATCTCATGTTAAGTGAAGAGGTTTCCAACT 2399  
Qy 2281 GAAAGATCTGTAGAGATGACGATATTTCATTTGTTAGTGTGAGAAAGGTTTCCAACT 2340  
Db 2400 GAAAGATCTGTAGAGATGACGATATTTCATTTGTTAGTGTGAGAAAGGTTTCCAACT 2459  
Qy 2341 GAAAGTATCTGCTTACTGAGAGTGAAGTGAACCTTACGAGGAAAGCAAGAACCAATTA 2400  
Db 2460 GAAAGTATCTGCTTACTGAGAGTGAAGTGAACCTTACGAGGAAAGCAAGAACCAATTA 2519  
Qy 2401 TGTGTAGTCAAGTGTGACATTTGAAACCCCAAGGACTAATTCATGTTGTTCCAA 2460  
Db 2520 TGTGTAGTCAAGTGTGACATTTGAAACCCCAAGGACTAATTCATGTTGTTCCAA 2579  
Qy 2461 GATATATGAATATGACAGAGAGGCTTAAGTATTCATTTGGACATGAAGTTAACACAGT 2520  
Db 2580 GATATATGAATATGACAGAGAGGCTTAAGTATTCATTTGGACATGAAGTTAACACAGT 2639  
Qy 2521 CGGGAACAGCATAGAAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580  
Db 2640 CGGGAACAGCATAGAAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2699  
Qy 2581 TTTCAAGTTCCTCAAGCCGACATCTGCTGTTTCAATTCAGAGAAATGCGAAGAG 2640  
Db 2700 TTTCAAGTTCCTCAAGCCGACATCTGCTGTTTCAATTCAGAGAAATGCGAAGAG 2759  
Qy 2641 GAATGTCAACATTTCTGCCCACCTGCTGCTTTAAAGAAACAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTCAACATTTCTGCCCACCTGCTGCTTTAAAGAAACAAGTCCAAAGTCACT 2819  
Qy 2701 TTTGAATGTGAACAAAGAGAAAGAAATCAAGGAAAGATGATTAATATCAACCTGTA 2760  
Db 2820 TTTGAATGTGAACAAAGAGAAAGAAATCAAGGAAAGATGATTAATATCAACCTGTA 2879  
Qy 2761 CAGACATTAATATCACTGACAGGCTTCTGCTGTTGTCAGAAAGTAAAGCCAGTTGAT 2820  
Db 2880 CAGACATTAATATCACTGACAGGCTTCTGCTGTTGTCAGAAAGTAAAGCCAGTTGAT 2939  
Qy 2821 AATGCCAATGTAGTATCAAGAGAGGCTAGGTTTGTATCATCTCACTTACAGAGC 2880  
Db 2940 AATGCCAATGTAGTATCAAGAGAGGCTAGGTTTGTATCATCTCACTTACAGAGC 2999  
Qy 2881 AAGCAACTGGACATTAATCCAAATTAACATGAGCTTTTACAAACCCATATCGATA 2940  
Db 3000 AAGCAACTGGACATTAATCCAAATTAACATGAGCTTTTACAAACCCATATCGATA 3059  
Qy 2941 CCACACATTTTCCCATCAAGTCAATTTGTTAAACCTAAATGTAAGAAATCTCTGAG 3000  
Db 3060 CCACACATTTTCCCATCAAGTCAATTTGTTAAACCTAAATGTAAGAAATCTCTGAG 3119  
Qy 3001 GAAAACTTTGAGGAACATTAATGTACCTGAAAGAGAAATGGGAAATGAGAACATTCCA 3060  
Db 3120 GAAAACTTTGAGGAACATTAATGTACCTGAAAGAGAAATGGGAAATGAGAACATTCCA 3179  
Qy 3061 AGTACATGAGCACAATTAACCGTAAATTAACATTTAGAGAAATGTTTAAAGAACGAGC 3120  
Db 3180 AGTACATGAGCACAATTAACCGTAAATTAACATTTAGAGAAATGTTTAAAGAACGAGC 3239

QY 3121 TCAAGCAATATTAAATAGTAGTTCACAGTACTAATGAATGGGCTCCAGTATTATGA 3180  
|||||  
Db 3240 TCAGCAATATTATTAAGTAGTTCACAGTACTAATGAATGGGCTCCAGTATTATGA 3299  
QY 3181 ATAGGTTCCAGTATGATAAATCAATTCAGAGCACTAGTATGAATGAGAGGCCAAATG 3240  
|||||  
Db 3300 ATAGGTTCCAGTATGATAAATCAATTCAGAGCACTAGTATGAATGAGAGGCCAAATG 3359  
QY 3241 AATGCTATGCTAGATTAGGGGTTTTCAGCACTGAGTCTATTAACAAAGCTTCCGGA 3300  
|||||  
Db 3360 AATGCTATGCTAGATTAGGGGTTTTCAGCACTGAGTCTATTAACAAAGCTTCCGGA 3419  
QY 3301 ACTAATTTGAAGCATCTGTAATTAAGCAAGAAATATGAAGAGTAGTTGAGACTGT 3360  
|||||  
Db 3420 ACTAATTTGAAGCATCTGTAATTAAGCAAGAAATATGAAGAGTAGTTGAGACTGT 3479  
QY 3361 AATPACGATTTCTCCATATCTGATTTGATTAATTTGAACAGCTATGGGAGTAGT 3420  
|||||  
Db 3480 AATPACGATTTCTCCATATCTGATTTGATTAATTTGAACAGCTATGGGAGTAGT 3539  
QY 3421 CATGCTATCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGGTGAATTAAG 3480  
|||||  
Db 3540 CATGCTATCAGGTTTGTCTGAGACACCTGATGACCTGTAGATGATGGTGAATTAAG 3599  
QY 3481 GAAGATACTAGTTTGTCTGAAAAATGACATTTAAGGAAAGTTCTGCTGTTTACCAAAAGC 3540  
|||||  
Db 3600 GAAGATACTAGTTTGTCTGAAAAATGACATTTAAGGAAAGTTCTGCTGTTTACCAAAAGC 3659  
QY 3541 GTCAGAAAGAGAGCTTACGAGAGTCCAGTCCGATTCACCATACACTTTGGGCTCAG 3600  
|||||  
Db 3660 GTCAGAAAGAGAGCTTACGAGAGTCCAGTCCGATTCACCATACACTTTGGGCTCAG 3719  
QY 3601 GGTACCGAAGAGGGGCGCAAGAAATTAAGTCTCTGAGAGAGAACTATCTAGTAGAGAT 3660  
|||||  
Db 3720 GGTACCGAAGAGGGGCGCAAGAAATTAAGTCTCTGAGAGAGAACTATCTAGTAGAGAT 3779  
QY 3661 GAAGAGCTTCCCTGCTTCCAACTTGTATTGTTGTTAAGTAACAAATATACCTTCTCAG 3720  
|||||  
Db 3780 GAAGAGCTTCCCTGCTTCCAACTTGTATTGTTGTTAAGTAACAAATATACCTTCTCAG 3839  
QY 3721 TCTACTAGGATAGACCGTGTGCTACGAGTGTCTGTAGAAACAGAGGAGAAATTA 3780  
|||||  
Db 3840 TCTACTAGGATAGACCGTGTGCTACGAGTGTCTGTAGAAACAGAGGAGAAATTA 3899  
QY 3781 TTATCATTTGAAGAAATAGCTTAATTAAGTCTGAGTAACAGGATTAATTTGGCAAGGCAATCT 3840  
|||||  
Db 3900 TTATCATTTGAAGAAATAGCTTAATTAAGTCTGAGTAACAGGATTAATTTGGCAAGGCAATCT 3959  
QY 3841 CAGGAACATCACCTTAGTATGAGGAAACAAATGTTCTGCTAGCTGTTTCTTACAGTGC 3900  
|||||  
Db 3960 CAGGAACATCACCTTAGTATGAGGAAACAAATGTTCTGCTAGCTGTTTCTTACAGTGC 4019  
QY 3901 AGTGAATTTGGAAGACTTGACTGCAATTAACAAACAGGAGATCCTTCTGATTGGTTC 3960  
|||||  
Db 4020 AGTGAATTTGGAAGACTTGACTGCAATTAACAAACAGGAGATCCTTCTGATTGGTTC 4079  
QY 3961 TCCAAACAATAGAGCATAGTCTGAAGCCAGAGGAGTTGCTGAGTGAACAAGAAATG 4020  
|||||  
Db 4080 TCCAAACAATAGAGCATAGTCTGAAGCCAGAGGAGTTGCTGAGTGAACAAGAAATG 4139  
QY 4021 GTTTAGATGATGAAGAAAGAGAGACGGCTTGAAGAAATATTAACAAGAGCAAGC 4080  
|||||  
Db 4140 GTTTAGATGATGAAGAAAGAGAGACGGCTTGAAGAAATATTAACAAGAGCAAGC 4199  
QY 4081 ATGATTTCAACTTAGTGAAGAGCATGGGTGAGAGTGAACAAGCGCTCTGAA 4140  
|||||  
Db 4200 ATGATTTCAACTTAGTGAAGAGCATGGGTGAGAGTGAACAAGCGCTCTGAA 4259  
QY 4141 GATGCTCAGGGCTATCTCTGAGAGTGAATTTTAACCACTCAGAGAGGAGTACATG 4200  
|||||  
Db 4260 GATGCTCAGGGCTATCTCTGAGAGTGAATTTTAACCACTCAGAGAGGAGTACATG 4319  
QY 4201 CAACATTAACCTGATAAGCTCCAGCAGAGAAATGGCTGAACCTGTGTTAGAACAG 4260

Db 4320 CAACATTAACCTGATAAGCTCCAGCAGAGAAATGGCTGAACCTAGAGCTGTGTTAGAACAG 4379  
QY 4261 CATGGAGCGCAGCTTCTTAACAGCTACCTTCATCATATAGTACCTTCTGCTCCTAG 4320  
Db 4380 CATGGAGCGCAGCTTCTTAACAGCTACCTTCATCATATAGTACCTTCTGCTCCTAG 4439  
QY 4321 GACCTGCAAAATCCAGAACAAAGACATCAGAAAAAGCAGTATTAATCTTACAGAAAAAGT 4380  
|||||  
Db 4440 GACCTGCAAAATCCAGAACAAAGACATCAGAAAAAGCAGTATTAATCTTACAGAAAAAGT 4499  
QY 4381 AGTGAATTAACCTTAATGAAGCAATCCAGAGGCTTCTGCTGACAAAGTTGAGTGTCT 4440  
|||||  
Db 4500 AGTGAATTAACCTTAATGAAGCAATCCAGAGGCTTCTGCTGACAAAGTTGAGTGTCT 4559  
QY 4441 GCAGATAGTCTACCGATTAATAAGAACAGGAGTGAAGAGGATCCCTCTPAAA 4500  
|||||  
Db 4560 GCAGATAGTCTACCGATTAATAAGAACAGGAGTGAAGAGGATCCCTCTPAAA 4619  
QY 4501 TGCCCATCATTAAGATAGTGTGTACATGACAGAGTTGCTTGGAGTCTTCAGAAATGA 4560  
|||||  
Db 4620 TGCCCATCATTAAGATAGTGTGTACATGACAGAGTTGCTTGGAGTCTTCAGAAATGA 4679  
QY 4561 AACTACCATCTCAAGAGAGCTCATTAAGGTTGTTGATGTGAGAGACACAGCTGGA 4620  
|||||  
Db 4680 AACTACCATCTCAAGAGAGAGCTCATTAAGGTTGTTGATGTGAGAGACACAGCTGGA 4739  
QY 4621 GAGTCTGGGCGCACAGATTTGAAGGAAACATCTTCTCCAAAGGCAAGATCTAGAGGA 4680  
|||||  
Db 4740 GAGTCTGGGCGCACAGATTTGAAGGAAACATCTTCTCCAAAGGCAAGATCTAGAGGA 4799  
QY 4681 ACCCTTACCTGGAATCTGGAATTCAGCCCTTCTCTGATGACCCGTAATCTGATCCTCT 4740  
|||||  
Db 4800 ACCCTTACCTGGAATCTGGAATTCAGCCCTTCTCTGATGACCCGTAATCTGATCCTCT 4859  
QY 4741 GAAGACAGAGCCCGAGAGTACGCTGTGTGGCAACATACCTTCTCAACCTCTGATG 4800  
|||||  
Db 4860 GAAGACAGAGCCCGAGAGTACGCTGTGTGGCAACATACCTTCTCAACCTCTGATG 4919  
QY 4801 AAGGTTCCCAATTTGAAGTTGAGAAATCTGGCCAGAGTCCAGTGTGCTCATCTACT 4860  
|||||  
Db 4920 AAGGTTCCCAATTTGAAGTTGAGAAATCTGGCCAGAGTCCAGTGTGCTCATCTACT 4979  
QY 4861 GATACGTCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAGACCAAGAAATGACA 4920  
|||||  
Db 4980 GATACGTCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGGAGAGACCAAGAAATGACA 5039  
QY 4921 GCTTCAACAGAAAGGTTCAACAAAGAAATGTCCATGTGTGTCTGGCTGACCCAGAA 4980  
|||||  
Db 5040 GCTTCAACAGAAAGGTTCAACAAAGAAATGTCCATGTGTGTCTGGCTGACCCAGAA 5099  
QY 4981 GAATTTATCTGCTGTGACAAAGTTTGCAGAAACACCAATCATCTTAACTAATTAAT 5040  
|||||  
Db 5100 GAATTTATCTGCTGTGACAAAGTTTGCAGAAACACCAATCATCTTAACTAATTAAT 5159  
QY 5041 ACTGAAGAGCTACTCATGTTGTTAAGAAACAGATGCTGAGTTTGTGTGAACGAGCA 5100  
|||||  
Db 5160 ACTGAAGAGCTACTCATGTTGTTAAGAAACAGATGCTGAGTTTGTGTGAACGAGCA 5219  
QY 5101 CTGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTTAGTATTTCTGGGTGACC 5160  
|||||  
Db 5220 CTGAATATTTTCTAGGAATTTGGGAGAGAAATGGGTAGTTAGTATTTCTGGGTGACC 5279  
QY 5161 CAGTCTATTAAGAAAGAAATGCTGAATGAGATGATTTTGAATCAGAGGAGATG 5220  
|||||  
Db 5280 CAGTCTATTAAGAAAGAAATGCTGAATGAGATGATTTTGAATCAGAGGAGATG 5339  
QY 5221 GTCAATGGAAGAACCCAGAGTCCAAAGCAGAGCAAGATCCAGAGCAAGAAATG 5280  
|||||  
Db 5340 GTCAATGGAAGAACCCAGAGTCCAAAGCAGAGCAAGATCCAGAGCAAGAAATG 5399  
QY 5281 TTCAGGGGCTGGAATCTGTTGCTATGGGCCCTTACCAACATGCTCCACAGATCACTG 5340  
|||||



Db 5400 TTCAGGGGGCTAGAAATCTGTGCTATGGGCCCTTCACCAACATGCCACAGATCAACTG 5459  
Qy 5441 GAATGATGATGACAGCTGTGGTCTGTGGTGAAGAGCTTCATATTCACCTT 5400  
Db 5460 GAATGATGATGACAGCTGTGGTCTGTGGTGAAGAGCTTCATATTCACCTT 5519  
Qy 5401 GGCACAGGTGTCCACCCCAATTTGGTGTGTGACAGCCAGATGCTTGACAGAGCAATGGC 5460  
Db 5520 GGCACAGGTGTCCACCCCAATTTGGTGTGTGACAGCCAGATGCTTGACAGAGCAATGGC 5579  
Qy 5461 TTCATGCAATTTGGGCGAGATGTGTGAGGCACTGTGTGACCCGAGAGTGGTGTGGAC 5520  
Db 5580 TTCATGCAATTTGGGCGAGATGTGTGAGGCACTGTGTGACCCGAGAGTGGTGTGGAC 5639  
Qy 5521 AGTGTAGACTCTACCAATGCGCAGAGCTGGACACCTGATATCCCGATCCGCCAC 5580  
Db 5640 AGTGTAGACTCTACCAATGCGCAGAGCTGGACACCTGATATCCCGATCCGCCAC 5699  
Qy 5581 AGCCACTAC 5589  
Db 5700 AGCCACTAC 5708

RESULT 12  
176943 176943 5914 bp DNA Linear PAT 03-APR-1998  
LOCUS Sequence 1 from patent US 5693473.  
ACCESSION 176943  
VERSION 176943.1 GI:3013097  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5914)  
AUTHORS Shattuck-Eidens,D.M., Simard,J., Durocher,F., Eml,M. and Nakamura,Y.  
TITLE Linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5693473-A 1 02-DEC-1997;  
FEATURES  
source location/Qualifiers  
1..5914  
BASE COUNT 2006 a 1156 c 1316 g 1436 t  
ORIGIN

Query Match 99.9%; Score 5585.8; DB 6; Length 5914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGATTTTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTAATGCTATGCAGAAA 60  
Db 120 ATGATTTTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTAATGCTATGCAGAAA 179  
Qy 61 ATCTTGAAGTGTCCCATCTGTCTGAGAGTATCAAGAACCTGTCTCCACAAAGTGTGAC 120  
Db 180 ATCTTGAAGTGTCCCATCTGTCTGAGAGTATCAAGAACCTGTCTCCACAAAGTGTGAC 239  
Qy 121 CACATATTTTGGCAATTTTGCATGCTGAACCTCTCAACCAAGAAAGAGGCTTTCACAG 180  
Db 240 CACATATTTTGGCAATTTTGCATGCTGAACCTCTCAACCAAGAAAGAGGCTTTCACAG 299  
Qy 181 TGTCTTATATGATGAATGATATACCAAAAGAGCCTTACAGAAATAGAGATTTAGT 240  
Db 300 TGTCTTATATGATGAATGATATACCAAAAGAGCCTTACAGAAATAGAGATTTAGT 359  
Qy 241 CAATCTTTGAAGAGCTATTGAAATCATTTTGTCTTTTACGCTTGACACAGGTTTGGAG 300  
Db 360 CAATCTTTGAAGAGCTATTGAAATCATTTTGTCTTTTACGCTTGACACAGGTTTGGAG 419  
Qy 301 TATGCAAAAGAGCTATATTTTGAAGAAAGAAATATCTCTCTGAACATCTTAAAGAT 360  
Db 420 TATGCAAAAGAGCTATATTTTGAAGAAAGAAATATCTCTCTGAACATCTTAAAGAT 479  
Qy 361 GAATTTCTATCATCCAAAGTATGGGCTACAGAAACCGTCCAAAGACTTCTACAGAGT 420

Db 480 GAATTTCTATCATCCAAAGATATGGGCTACAGAAACCGTGCAGAAAGACTTCTACAGAGT 539  
Qy 421 GAACCCGAAATTCCTTCTTGCAGAAACCAAGTCTCAGTGTCCAACTCTCTAACCTTGA 480  
Db 540 GAACCCGAAATTCCTTCTTGCAGAAACCAAGTCTCAGTGTCCAACTCTCTAACCTTGA 599  
Qy 481 ACTGTGAAGACTCTGAGAGCAACGCGGATACAACTCAAAAAGAGCTGTCTACATTT 540  
Db 600 ACTGTGAAGACTCTGAGAGCAACGCGGATACAACTCAAAAAGAGCTGTCTACATTT 659  
Qy 541 GAATTTGGATCTGATTTCTTCTGAAAGATATCCGTTAATTAAGCCAACTTATTCAGTGTGGA 600  
Db 660 GAATTTGGATCTGATTTCTTCTGAAAGATATCCGTTAATTAAGCCAACTTATTCAGTGTGGA 719  
Qy 601 GATCAGCAATTTGTTACAATTCACCCCTCAAGAAACCAAGGATGAATTCAGTGTGATCT 660  
Db 720 GATCAGCAATTTGTTACAATTCACCCCTCAAGAAACCAAGGATGAATTCAGTGTGATCT 779  
Qy 661 GCAGAAAGAGCTGCTGTGTAATTTCTGAGACGAGATGTAACAAATACGAAATCATCA 720  
Db 780 GCAGAAAGAGCTGCTGTGTAATTTCTGAGACGAGATGTAACAAATACGAAATCATCA 839  
Qy 721 CCCAGTAAATGATTTTGAACACACCTGAGAAAGCGTGCAGCTGAGAGCATCCAGAAAG 780  
Db 840 CCCAGTAAATGATTTTGAACACACCTGAGAAAGCGTGCAGCTGAGAGCATCCAGAAAG 899  
Qy 781 TATCAGGGATAGTCTGTTTCAAACTGATGTGAGCCATGTGGCACAAATACTCATGCC 840  
Db 900 TATCAGGGATAGTCTGTTTCAAACTGATGTGAGCCATGTGGCACAAATACTCATGCC 959  
Qy 841 AGCTCATTAAGATGAGAACAGAGCTTATTAATCTCAATAAGACAAATGAATGATGA 900  
Db 960 AGCTCATTAAGATGAGAACAGAGCTTATTAATCTCAATAAGACAAATGAATGATGA 1019  
Qy 901 AAGGCTAAATTTCTGTAATTAAGCAACACCTGCTTACCAAGAGCACCACATACAGA 960  
Db 1020 AAGGCTAAATTTCTGTAATTAAGCAACACCTGCTTACCAAGAGCACCACATACAGA 1079  
Qy 961 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACCTCCACACAGAAAAAGGTA 1020  
Db 1080 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACCTCCACACAGAAAAAGGTA 1139  
Qy 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATAGCAAAATGCCATGC 1080  
Db 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATAGCAAAATGCCATGC 1199  
Qy 1081 TCAGAGAAATCTAGAGATGATGAAAGATGTTCTTGGATTAACACTAAATAGCAGCATTCAG 1140  
Db 1200 TCAGAGAAATCTAGAGATGATGAAAGATGTTCTTGGATTAACACTAAATAGCAGCATTCAG 1259  
Qy 1141 AAGGTAATGAGTGTGTTTCCAGAGTGAATGACTGTTAGTGTGATGACTCATGAT 1200  
Db 1260 AAGGTAATGAGTGTGTTTCCAGAGTGAATGACTGTTAGTGTGATGACTCATGAT 1319  
Qy 1201 GGGGAGTCAATCAATATGCAAGTATGATGATATTTGACGTTCAATAGAGCTGAT 1260  
Db 1320 GGGGAGTCAATCAATATGCAAGTATGATGATATTTGACGTTCAATAGAGCTGAT 1379  
Qy 1261 GAATATTTCTGTTCTTACAGAGAAATAGACTTACTGSCAGTACTCTCATGAGGCTTTA 1320  
Db 1380 GAATATTTCTGTTCTTACAGAGAAATAGACTTACTGSCAGTACTCTCATGAGGCTTTA 1439  
Qy 1321 ATATGTAAGTGAAGAGTTCCTCAATCAATCAGTAGAGATTAATTTGAAGACAAATA 1380  
Db 1440 ATATGTAAGTGAAGAGTTCCTCAATCAATCAGTAGAGATTAATTTGAAGACAAATA 1499  
Qy 1381 TTTGGGAAACCTATGGAAGAGGCAAGGCTCCCAACTTAAGCCATGTAACGAAAT 1440  
Db 1500 TTTGGGAAACCTATGGAAGAGGCAAGGCTCCCAACTTAAGCCATGTAACGAAAT 1559  
Qy 1441 CTAAATATAGAGCATTTGTTACTGAGCCACAGATAATATCAAGAGCTCCCTCACAAT 1500

Db 1560 CTAATATAGGACATTTGTTACTGAGCCACAGATAATACAGAGCGTCCCTCACAAAT 1619  
QY 1501 AATTTAAAGGCTAAAGAGACCTAGATCAGGCTTCATCTGAGATTTTATCAAGAA 1560  
Db 1620 AATTTAAAGGCTAAAGAGACCTAGATCAGGCTTCATCTGAGATTTTATCAAGAA 1619  
QY 1561 GCAGATTTGGCAGTTCAAAAGACTCCTGAATGATAATCAGGAACTAACCAACGGAG 1620  
Db 1680 GCAGATTTGGCAGTTCAAAAGACTCCTGAATGATAATCAGGAACTAACCAACGGAG 1739  
QY 1621 CAGAAATGCTCAAGTGAATATTTACTAATAGTGGTCATGAGATTAACAAAAGTGAT 1680  
Db 1740 CAGAAATGCTCAAGTGAATATTTACTAATAGTGGTCATGAGATTAACAAAAGTGAT 1799  
QY 1681 TCTATTGAGATGAGAAAAATCCTTAACCAATAGATCACTCGAAAAAGATGCTGCTTC 1740  
Db 1800 TCTATTGAGATGAGAAAAATCCTTAACCAATAGATCACTCGAAAAAGATGCTGCTTC 1659  
QY 1741 AAAAGCAAGCTGACCTTATAGCAGCAGTATAGCAATATGGAACCTCGAATTAATATC 1800  
Db 1860 AAAAGCAAGCTGACCTTATAGCAGCAGTATAGCAATATGGAACCTCGAATTAATATC 1919  
QY 1801 CACAATTTCAAAAGCACCTTAAAAAGATAGGCTGAGGAGAGTCTTCTACAGGCAATAT 1860  
Db 1920 CACAATTTCAAAAGCACCTTAAAAAGATAGGCTGAGGAGAGTCTTCTACAGGCAATAT 1979  
QY 1861 CATGCGCTTGAACCTAGTACTAGTAAATCTAAGCCACCTAATTTGTAATTCGA 1920  
Db 1980 CATGCGCTTGAACCTAGTACTAGTAAATCTAAGCCACCTAATTTGTAATTCGA 2039  
QY 1921 ATTGATAGTGTCTGACAGTGAAGATTAAGAAAAAGTACAAACCAATGCCAGTC 1980  
Db 2040 ATTGATAGTGTCTGACAGTGAAGATTAAGAAAAAGTACAAACCAATGCCAGTC 2099  
QY 1981 AGGCAAGAGAGAAACCTTACAACTCATGAGAGTAAAGAACCTGCAACTGGAGCCAGAG 2040  
Db 2100 AGGCAAGAGAGAAACCTTACAACTCATGAGAGTAAAGAACCTGCAACTGGAGCCAGAG 2159  
QY 2041 AGTAAACAGCCAAATGACAGACAGTAAGAAAGCATGACAGTACTTTCCAGAGCTG 2100  
Db 2160 AGTAAACAGCCAAATGACAGACAGTAAGAAAGCATGACAGTACTTTCCAGAGCTG 2219  
QY 2101 AAGTTAAACAAATGACAGTGTCTTTTACTAAGTGTTCAAATACCACTGAACCTTAAGAA 2160  
Db 2220 AAGTTAAACAAATGACAGTGTCTTTTACTAAGTGTTCAAATACCACTGAACCTTAAGAA 2279  
QY 2161 TTTGTCATCTAGACCTTCCAAAGAGAGAAAAAGAGAGAACTAGAAACAGTTAAAGT 2220  
Db 2280 TTTGTCATCTAGACCTTCCAAAGAGAGAAAAAGAGAGAACTAGAAACAGTTAAAGT 2339  
QY 2221 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAGAAAGGTTTTGCAACT 2280  
Db 2340 TCTAATATGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAGAAAGGTTTTGCAACT 2399  
QY 2281 GAAATATCTGAGAGTGAAGTATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340  
Db 2400 GAAATATCTGAGAGTGAAGTATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2459  
QY 2341 GAAATATCTGCTTACTGGAAGTTAGACCTTGAAGAGAGCAAAAAAGAGAGAACTAA 2400  
Db 2460 GAAATATCTGCTTACTGGAAGTTAGACCTTGAAGAGAGCAAAAAAGAGAGAACTAA 2519  
QY 2401 TGTGTGAGTCAAGTGTGACACATTTGAAAAACCCCAAGGACTAATTCATGGTTGCTCA 2460  
Db 2520 TGTGTGAGTCAAGTGTGACACATTTGAAAAACCCCAAGGACTAATTCATGGTTGCTCA 2579  
QY 2461 GATTAATGAGATGACACAGAGGCTTTAAGTATCATTGGGACATGAAGTTAACCAAGT 2520  
Db 2580 GATTAATGAGATGACACAGAGGCTTTAAGTATCATTGGGACATGAAGTTAACCAAGT 2639  
QY 2521 GGGGAACAGAGATGAGAAATGAGAGAAAGCACTGATGCTCAGATTTGCAAGATPAC 2580  
Db 2640 GGGGAACAGAGATGAGAAATGAGAGAAAGCACTGATGCTCAGATTTGCAAGATPAC 2699

QY 2581 TTCAAGGTTTCAAGGCGCAGTCAATTTGCTGTGTTTCAATTCAGAGAAATGACAGAG 2640  
Db 2700 TTCAAGGTTTCAAGGCGCAGTCAATTTGCTGTGTTTCAATTCAGAGAAATGACAGAG 2759  
QY 2641 GAATGTGCAACATTTCTGTGCGCCACTGTGGTCTTTAAGAAACCAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGCAACATTTCTGTGCGCCACTGTGGTCTTTAAGAAACCAAGTCCAAAGTCACT 2819  
QY 2701 TTTGAATGGAACAAAAGAGAGAAATCAAGGAAAGAGTCAATATCAAGCTGTA 2760  
Db 2820 TTTGAATGGAACAAAAGAGAGAAATCAAGGAAAGAGTCAATATCAAGCTGTA 2879  
QY 2761 CAGACAGTAAATATCACTGAGGCTTCTGTGTTGCTGAGAGAGATTAAGCAGTGTAT 2820  
Db 2880 CAGACAGTAAATATCACTGAGGCTTCTGTGTTGCTGAGAGAGATTAAGCAGTGTAT 2939  
QY 2821 AATGCCAATGATGATTCAGAGAGGCTTACGTTTGTCTATCTCAGTTCAGAGGC 2880  
Db 2940 AATGCCAATGATGATTCAGAGAGGCTTACGTTTGTCTATCTCAGTTCAGAGGC 2999  
QY 2881 AAGCAACCTGAGCTCATTTACTCCAAATTAACATGACCTTTTCAAAACCCATATGTA 2940  
Db 3000 AAGCAACCTGAGCTCATTTACTCCAAATTAACATGACCTTTTCAAAACCCATATGTA 3059  
QY 2941 CCACACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3000  
Db 3060 CCACACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAG 3119  
QY 3001 GAAACTTTGAGAGAACTCAATGTCACCTGAAGAGAAATGGAATGAGAACCTTCCA 3060  
Db 3120 GAAACTTTGAGAGAACTCAATGTCACCTGAAGAGAAATGGAATGAGAACCTTCCA 3179  
QY 3061 AGTACAGTGAAGCAATTTAGCCGTATTAACATTAGAGAAAAAGTTTTAAAGAACCCAGC 3120  
Db 3180 AGTACAGTGAAGCAATTTAGCCGTATTAACATTAGAGAAAAAGTTTTAAAGAACCCAGC 3239  
QY 3121 TCAAGCAATATTAATGAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAAGAA 3180  
Db 3240 TCAAGCAATATTAATGAGTAGTTCAGTACTAATGAAGTGGCTCCAGTATTAAGAA 3299  
QY 3181 ATAGTGTCCAGATGATGAAACATTCAGAGAGAACTAGAGAAAGAGGGCCCAAAATG 3240  
Db 3300 ATAGTGTCCAGATGATGAAACATTCAGAGAGAACTAGAGAAAGAGGGCCCAAAATG 3359  
QY 3241 AATGCTATCTTATAGTTAGGGTTTGGCAACTGAGGCTATTAAGCAAAAGTCTTCTGGA 3300  
Db 3360 AATGCTATCTTATAGTTAGGGTTTGGCAACTGAGGCTATTAAGCAAAAGTCTTCTGGA 3419  
QY 3301 AGTAAATGTAAGCATCTGAATTAAGAAAGCAAGAAATGAGAAAGTTCAGACTGTT 3360  
Db 3420 AGTAAATGTAAGCATCTGAATTAAGAAAGCAAGAAATGAGAAAGTTCAGACTGTT 3479  
QY 3361 AATACAGATTTCTCTCATATGATTTGAGATTAACCTTGAACACCTATGGAAGTATG 3420  
Db 3480 AATACAGATTTCTCTCATATGATTTGAGATTAACCTTGAACACCTATGGAAGTATG 3539  
QY 3421 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTGATGATGAGTGAATAAG 3480  
Db 3540 CATGCACTCAGGTTTGTCTGAGACACCTGATGACCTGTTGATGATGAGTGAATAAG 3599  
QY 3481 GAAGATTAAGTATTTGCTGAAAATGACATTTAAGAAAGTTCGCTGTTTTAGCAAAAGC 3540  
Db 3600 GAAGATTAAGTATTTGCTGAAAATGACATTTAAGAAAGTTCGCTGTTTTAGCAAAAGC 3659  
QY 3541 GTCCAGAAAGAGAGCTTATAGCAGAGCTCTAGACCTTTACACCATATTCATTTGGCTCAG 3600  
Db 3660 GTCCAGAAAGAGAGCTTATAGCAGAGCTCTAGACCTTTACACCATATTCATTTGGCTCAG 3719  
QY 3601 GGTACCGAAGAGGGCCCAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT 3660  
Db 3720 GGTACCGAAGAGGGCCCAAGAAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT 3779



TITLE Linked breast and ovarian cancer susceptibility gene  
JOURNAL Patent: US 5709999-A 1 20-JAN-1998:  
FEATURES Location/Qualifiers  
Source 1. 5914  
/organism="unknown"  
BASE COUNT 2006 a 1156 c 1316 g 1436 t  
ORIGIN

Query Match 99.9%; Score 5585.8; DB 6; Length 5914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGATTTATCTGCTCTTGGCGTTGAAGAATGATTAATGCTATGAGAAA 60  
DB 120 ATGATTTATCTGCTCTTGGCGTTGAAGAATGATTAATGCTATGAGAAA 179  
QY 61 ATCTAGAGTGTCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCAAAGTGTAC 120  
DB 180 ATCTAGAGTGTCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCAAAGTGTAC 239  
QY 121 CACATATTTTGCATAATTTTGCATGCTGAACCTCTCAACGAGAAAGAGGCTTCACAG 180  
DB 240 CACATATTTTGCATAATTTTGCATGCTGAACCTCTCAACGAGAAAGAGGCTTCACAG 299  
QY 181 TGTCTTTATGTAAGATGATTAACCAAAGAGGCTCAAGAAAGTACGAGATTAGT 240  
DB 300 TGTCTTTATGTAAGATGATTAACCAAAGAGGCTCAAGAAAGTACGAGATTAGT 359  
QY 241 CAACCTGTTGAGAGCTATGAAATCATTGTGCTTTACAGCTTACACAGTTTGGAG 300  
DB 360 CAACCTGTTGAGAGCTATGAAATCATTGTGCTTTACAGCTTACACAGTTTGGAG 419  
QY 301 TATGCAACAGCTATATATTTTGCATAATTTTGCATGCTGAACATCTAAAGAT 360  
DB 420 TATGCAACAGCTATATATTTTGCATAATTTTGCATGCTGAACATCTAAAGAT 479  
QY 361 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCAGAAAGACTTCTACAGAT 420  
DB 480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCAGAAAGACTTCTACAGAT 539  
QY 421 GAACCCGAAATCCTTCTGCAAGAACCAAGTCTAGTCCAACTCTTAACCTTGA 480  
DB 540 GAACCCGAAATCCTTCTGCAAGAACCAAGTCTAGTCCAACTCTTAACCTTGA 599  
QY 481 ACTGTGAGACCTGTGAGGACAAAGAGCGGATACAACTCAAAAGAGCTGTCTACAT 540  
DB 600 ACTGTGAGACCTGTGAGGACAAAGAGCGGATACAACTCAAAAGAGCTGTCTACAT 659  
QY 541 GAATTTGGATCTGATTTCTTGAGATACCGTTAATTAAGGCACTTATTCAGTGTGGA 600  
DB 660 GAATTTGGATCTGATTTCTTGAGATACCGTTAATTAAGGCACTTATTCAGTGTGGA 719  
QY 601 GATCAAGAAATGTTCAATCAATCACCCTCAAGGACAGGATGAATCGATTGATTT 660  
DB 720 GATCAAGAAATGTTCAATCAATCACCCTCAAGGACAGGATGAATCGATTGATTT 779  
QY 661 GCAAAAAAGGCTGCTTGAATTTCTGAGAGGATGTAAACAATTAATCAATCATCAA 720  
DB 780 GCAAAAAAGGCTGCTTGAATTTCTGAGAGGATGTAAACAATTAATCAATCATCAA 839  
QY 721 CCCAGTAATTAATGATTGAACACCACTGAGAAAGCGTGCAGTGAAGGATCCAGAAAG 780  
DB 840 CCCAGTAATTAATGATTGAACACCACTGAGAAAGCGTGCAGTGAAGGATCCAGAAAG 899  
QY 781 TATAGAGGTAGTCTGTTCAACTGATGTGAGGCAATGAGCAATTAATCTCATGCC 840  
DB 900 TATAGAGGTAGTCTGTTCAACTGATGTGAGGCAATGAGCAATTAATCTCATGCC 959  
QY 841 AGCTCATTAACAGCATGAGACAGAGTTTATTAATCACTAAAGACAGATGAATGTAGA 900  
DB 960 AGCTCATTAACAGCATGAGACAGAGTTTATTAATCACTAAAGACAGATGAATGTAGA 1019  
QY 901 AAGGCTGAATTTCTGTATTAAGCAAAAGCGCTGCTTGAAGAGGCAACATTAACAGA 960

DB 1020 AAGGCTGAATTTCTGTATTAAGCAAAAGCGCTGCTTGAAGAGGCAACATTAACAGA 1079  
QY 961 TGGGCTGGAAGTATAGAAACATGTATGATAGCGGACTCCAGACAGAAAAAGGTA 1020  
DB 1080 TGGGCTGGAAGTATAGAAACATGTATGATAGCGGACTCCAGACAGAAAAAGGTA 1139  
QY 1021 GATCTGAATCTGATCCCTGTGTGAGAGAAAGATGAATTAACCAAACTGCAATGC 1080  
DB 1140 GATCTGAATCTGATCCCTGTGTGAGAGAAAGATGAATTAACCAAACTGCAATGC 1139  
QY 1081 TCAGAGAACTCAGAGATTAAGATGCTTCTGATTAACCTAAATAGCAGATTAC 1140  
DB 1200 TCAGAGAACTCAGAGATTAAGATGCTTCTGATTAACCTAAATAGCAGATTAC 1259  
QY 1141 AAAGTTAATGAGTGGTTTCCAGAGAGTATGACACTGTTAGTTCGATGACTCAGAT 1200  
DB 1260 AAAGTTAATGAGTGGTTTCCAGAGAGTATGACACTGTTAGTTCGATGACTCAGAT 1319  
QY 1201 GGGAGTCTGAATCAAAATGCCAAAGTACGTATGATTGAGAGTTCTAAATGAGTAGAT 1260  
DB 1320 GGGAGTCTGAATCAAAATGCCAAAGTACGTATGATTGAGAGTTCTAAATGAGTAGAT 1379  
QY 1261 GAATATCTGTTCTTCAGAGAAATAGACTTACTGGCCAGTGAATCCTCATGAGCTTTA 1320  
DB 1380 GAATATCTGTTCTTCAGAGAAATAGACTTACTGGCCAGTGAATCCTCATGAGCTTTA 1439  
QY 1321 ATATGTAAAGTGAAGAGTTCACCTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1380  
DB 1440 ATATGTAAAGTGAAGAGTTCACCTCCAAATCAGTAGAGATTAATTTGAAGACAAATA 1499  
QY 1381 TTTGGGAAACCTATTCGAGAGAGCAAGCGCTCCCACTTAAGCATGTAACTGAAT 1440  
DB 1500 TTTGGGAAACCTATTCGAGAGAGCAAGCGCTCCCACTTAAGCATGTAACTGAAT 1559  
QY 1441 CTAAATTAATGAGCATTTTGTACTGAGCCACAGATAATTAAGAGCGTCCCTCACAAT 1500  
DB 1560 CTAAATTAATGAGCATTTTGTACTGAGCCACAGATAATTAAGAGCGTCCCTCACAAT 1619  
QY 1501 AAATTAAGACCTTAAGAGAGCACTACATACAGCCCTTCCTCGAGAGATTTATCAAAA 1560  
DB 1620 AAATTAAGACCTTAAGAGAGCACTACATACAGCCCTTCCTCGAGAGATTTATCAAAA 1679  
QY 1561 GCAGATTTGGGAGTTCAAAGAGCTCTGAATGATTAATTCAGGAGAACTAACCAAGCGAG 1620  
DB 1680 GCAGATTTGGGAGTTCAAAGAGCTCTGAATGATTAATTCAGGAGAACTAACCAAGCGAG 1739  
QY 1621 CAGAATGCTCAAGTATGATTAATTAATAGTGTCTATGAGAAATTAAGCAAAAGTGTAT 1680  
DB 1740 CAGAATGCTCAAGTATGATTAATTAATAGTGTCTATGAGAAATTAAGCAAAAGTGTAT 1799  
QY 1681 TCTATTCGAATGAGAAAAATCCTAACCAATAGAAATCCTGCAAAAAAGATCTGCTTC 1740  
DB 1800 TCTATTCGAATGAGAAAAATCCTAACCAATAGAAATCCTGCAAAAAAGATCTGCTTC 1859  
QY 1741 AAAAGCAAGAGCTGAATTAAGAGAGCTATTAAGCAATTAAGCAATTAAGCAATTAATC 1800  
DB 1860 AAAAGCAAGAGCTGAATTAAGAGAGCTATTAAGCAATTAAGCAATTAAGCAATTAATC 1919  
QY 1801 CACAATTAAGAGCACTTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAT 1860  
DB 1920 CACAATTAAGAGCACTTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAT 1979  
QY 1861 CATGGCTTGAACCTGATGATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 1920  
DB 1980 CATGGCTTGAACCTGATGATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 2039  
QY 1921 ATTGATAGTGTCTTACAGAGTGAAGATAAGAAAAAGTCAACCAAAATGCGAGTC 1980  
DB 2040 ATTGATAGTGTCTTACAGAGTGAAGATAAGAAAAAGTCAACCAAAATGCGAGTC 2099  
QY 1981 AAGGACAGCAAAACCTTAACATCATGAGAGATTAAGAAACCTGCACTGAGGCAAGAG 2040

Db	2100	AGGCACAGCAAAACCTACACACTGATGGAAGSTAAAGACCTGCAACTGGAGCCAAAGAC	2159
QY	2041	AGTAAACAACCCAAATGAAACAGACAGATGAAAAGACATGACAGTGTACTTCCAGAGCTG	2100
Db	2160	AGTAAACAAGCCAAATGAAACACACAAGTAAAAGACATGACAGCGATCTTCCAGAGCTG	2219
QY	2101	AAGTTAAACAATGCACTGGTCTTTTACTAAGTGTCAATTACCAGTGAACCTTAAAGAA	2160
Db	2220	AAGTTAAACAATGCACTGGTCTTTTACTAAGTGTCAATTACCAGTGAACCTTAAAGAA	2279
QY	2161	TTTGTGCAATCCTAGCCTTCGAAAGAGAAAGAAAAGAGAAACATGAAACAGTTAAAGTG	2220
Db	2280	TTTGTGCAATCCTAGCCTTCGAAAGAGAAAGAAAAGAGAAACATGAAACAGTTAAAGTG	2339
QY	2221	TCTATAATAGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGCAAACT	2280
Db	2340	TCTATAATAGCTGAGAGACCCCAAGATCTCATGTTAAGTGAGAAAGGGTTTGCAAACT	2399
QY	2281	GAAAGATCTGTAGAGAGTACGACATTTTCATTTGTTGTTACTGTTATGTCAGCTCAG	2340
Db	2400	GAAAGATCTGTAGAGAGTACGACATTTTCATTTGTTGTTACTGTTATGTCAGCTCAG	2459
QY	2341	GAAAGTATCTGCTTACTGGAAGTTAGACCTGTAGGSAAGGCAAAAACGAAACCAATPAA	2400
Db	2460	GAAAGTATCTGCTTACTGGAAGTTAGACCTGTAGGSAAGGCAAAAACGAAACCAATPAA	2519
QY	2401	TGTGTGACTAGTGTGACATTTTGAAACCACCAAGGACTAATTCATGTGTTCTCCAA	2460
Db	2520	TGTGTGACTAGTGTGACATTTTGAAACCACCAAGGACTAATTCATGTGTTCTCCAA	2579
QY	2461	GATATATGAAATGACACAGAAAGCCTTTTAAATATCCATTGGGACATGAAGTTAACACAGT	2520
Db	2580	GATATATGAAATGACACAGAAAGCCTTTTAAATATCCATTGGGACATGAAGTTAACACAGT	2639
QY	2521	CGGGAAACAAACATAGAAATGGAAGAAAGTAACCTGATGTCGATTTTGCAAAATACA	2580
Db	2640	CGGGAAACAAACATAGAAATGGAAGAAAGTAACCTGATGTCGATTTTGCAAAATACA	2699
QY	2581	TTCAAGGTTTCAAAAGCGCCAGTCAATTTGCTGTTTCAATCCAGGAAATGCGAGAAG	2640
Db	2700	TTCAAGGTTTCAAAAGCGCCAGTCAATTTGCTGTTTCAATCCAGGAAATGCGAGAAG	2759
QY	2641	GAAATGTGCAACATTTCTGCCCCACTGTGGTCTTTAAAGAAACAAGTCCAAAAGTCACT	2700
Db	2760	GAAATGTGCAACATTTCTGCCCCACTGTGGTCTTTAAAGAAACAAGTCCAAAAGTCACT	2819
QY	2701	TTTGAATGTGAACAAAAGGAAGAAATACAGGAAGAAATAGTATATCACGCTGTA	2760
Db	2820	TTTGAATGTGAACAAAAGGAAGAAATACAGGAAGAAATAGTATATCACGCTGTA	2879
QY	2761	CAGACAGTTAATATCACATGCAAGGCTTTCCTGTGTGTCGTAAGAAAGATTAAGCCAGTTGAT	2820
Db	2880	CAGACAGTTAATATCACATGCAAGGCTTTCCTGTGTGTGTCGTAAGAAAGATTAAGCCAGTTGAT	2939
QY	2821	AATGCCAAATGTATATCAAAAGAGAGGCTCTGAGTTTGTCTATCATCTCAGTTCAAGGC	2999
Db	2940	AATGCCAAATGTATATCAAAAGAGAGGCTCTGAGTTTGTCTATCATCTCAGTTCAAGGC	3059
QY	2881	AACGAACATGACATCATTAATCCCAATTAACATGAGACTTTTACAAACCCATATCGTATA	2940
Db	3000	AACGAACATGACATCATTAATCCCAATTAACATGAGACTTTTACAAACCCATATCGTATA	3059
QY	2941	CCACACACTTTTCCCATCAAGTCAATTTGTTTAAACTTAAATGTAGAAAAATCTGCTAGAG	3000
Db	3060	CCACACACTTTTCCCATCAAGTCAATTTGTTTAAACTTAAATGTAGAAAAATCTGCTAGAG	3119
QY	3001	GAAAACTTTGAGGAACATTCATGTGACCTGAAAGAGAAATGGGAAATGGAACATTCCA	3060
Db	3120	GAAAACTTTGAGGAACATTCATGTGACCTGAAAGAGAAATGGGAAATGGAACATTCCA	3179
QY	3061	AGTACAGTGAACAAATTAGCCGCTTAATACATTGAGAAAAATGTTTTTAAAGAACGACG	3120
Db	3180	AGTACAGTGAACAAATTAGCCGCTTAATACATTGAGAAAAATGTTTTTAAAGAACGACG	3239

OY	3121	TCACGACCAATATTAATGAGTACGGTTCCGCTACTATAAGAGGGGCTCCAGATTAAATGAA	3180
Db	3240	TCAACCAATATTAATGAAGTAGGTTCCGTAAGTACTAATGAAGAGGGCTCCAGATTAAATGAA	3299
OY	3181	ATAGGTTCCAGTGAATGAACCAATTCACGACCAACATAGTATGAAAAACAGAGGCCAAATATG	3240
Db	3300	ATAGGTTCCAGTGAATGAACCAATTCACGACCAACATAGTATGAAAAACAGAGGCCAAATATG	3359
OY	3241	AATGCTATGCTTACATTAGGGGTTTGCACCTGAGGCTATATAACAAAGTCCTTCGGA	3300
Db	3360	AATGCTATGCTTACATTAGGGGTTTGCACCTGAGGCTATATAACAAAGTCCTTCGGA	3419
OY	3301	AGTAATTGTAGCATCCCGAATATAAAAAGCAACAATATGAAGAAAGTATGACACTGTT	3360
Db	3420	AGTAATTGTAGCATCCCGAATATAAAAAGCAACAATATGAAGAAAGTATGACACTGTT	3479
OY	3361	AATACAGATTTCTCTCCATATCTGATATTCAGATTAACCTTAGAACACGCTATGGAAAGT	3420
Db	3480	AATACAGATTTCTCTCCATATCTGATATTCAGATTAACCTTAGAACACGCTATGGAAAGT	3539
OY	3421	CATGATCTCAGGTTTGTTGTCGACACACCTGATACCTGTTAGATGATGGTGAATTAAG	3480
Db	3540	CATGATCTCAGGTTTGTTGTCGACACCTGATACCTGTTAGATGATGGTGAATTAAG	3599
OY	3481	GAAGATATACTATTTGGCTGAAAAATGACATTAAGGAAAGTTCTGCTGTTTTTACCAAAAGC	3540
Db	3600	GAAGATATACTATTTGGCTGAAAAATGACATTAAGGAAAGTTCTGCTGTTTTTACCAAAAGC	3659
OY	3541	GTCCAGAAAGAGACGCTTAGCAGAGATCTTAGCCCTTTCACCCATACACATTTGGCTCAG	3600
Db	3660	GTCCAGAAAGAGAGCTTAGCAGAGATCTTAGCCCTTTCACCCATACACATTTGGCTCAG	3719
OY	3601	GTTTACCGAAGAGGGGCCAAGAAATTTAGATTCCTCACAAGAGAACTTATGAGAGAGAT	3660
Db	3720	GTTTACCGAAGAGGGGCCAAGAAATTTAGATTCCTCACAAGAGAACTTATGAGAGAGAT	3779
OY	3661	GAAAGACTTCCCTGCTTCCAAACACTTGTATTTTGGTAAAGTAAACAATATACCTTCAG	3720
Db	3780	GAAAGACTTCCCTGCTTCCAAACACTTGTATTTTGGTAAAGTAAACAATATACCTTCAG	3839
OY	3721	TCTACTAGGCCTAGCACCCGTTGCTACCGAGATGCTGTCTAAGAACACAGAGGAGAAATTTA	3780
Db	3840	TCTACTAGGCCTAGCACCCGTTGCTACCGAGATGCTGTCTAAGAACACAGAGGAGAAATTTA	3899
OY	3781	TTATCATTTGAAGATAGCTTAAATGACGTGCGATTAACAGAGTAATATGCGCAAAAGGCATCT	3840
Db	3900	TTATCATTTGAAGATAGCTTAAATGACGTGCGATTAACAGAGTAATATGCGCAAAAGGCATCT	3959
OY	3841	CAGGAACATCACCTTAGTAGGAGGAACAAATGTTCTGCTAGCTTGTCTTCACAGTGC	3900
Db	3960	CAGGAACATCACCTTAGTAGGAGGAACAAATGTTCTGCTAGCTTGTCTTCACAGTGC	4019
OY	3901	AGTAATTTGGAAGACTTGACTCGCAAAATATCAAAACACCCAGATCCCTTCTTGATTTGGTCT	3960
Db	4020	AGTAATTTGGAAGACTTGACTCGCAAAATATCAAAACACCCAGATCCCTTCTTGATTTGGTCT	4079
OY	3961	TCCAACAACATGAGGCATCACTGTGAAGACCCAGGAGTTGGTCTGAGTGAACAAGAAATG	4020
Db	4080	TCCAACAACATGAGGCATCACTGTGAAGACCCAGGAGTTGGTCTGAGTGAACAAGAAATG	4139
OY	4021	GTTTCAGATGTATGAAGAAAGAGGAGGAGGCTGTGGAAGAAATATATCAAGAAAGCAAGC	4080
Db	4140	GTTTCAGATGTATGAAGAAAGAGGAGGAGGCTGTGGAAGAAATATATCAAGAAAGCAAGC	4199
OY	4081	ATGGAATTCAAACTTAGGTGAAGACGACATCTGGGTGTGAGAGTGAACAAACGCTCTCGAA	4140
Db	4200	ATGGAATTCAAACTTAGGTGAAGACGACATCTGGGTGTGAGAGTGAACAAACGCTCTCGAA	4259
OY	4141	GACGTGCTCAGGGCTATCTCTCAGAGTGAACATTTTAAACCACTCAGACAGAGGATACATG	4200
Db	4260	GACGTGCTCAGGGCTATCTCTCAGAGTGAACATTTTAAACCACTCAGACAGAGGATACATG	4319

QY 4201 CAACATACTGATAAGCTCCAGCAGGAAATGGCTGAACGTAGAACCTGTGTAGAACG 4260  
 |||||  
 Db 4320 CAACATACTGATAAGCTCCAGCAGGAAATGGCTGAACGTAGAACCTGTGTAGAACG 4379  
 |||||  
 QY 4261 CATGGAGCCAGCCTTCTAACGCTACCTTCCATCATTAAGTACTCTTCTGCCCTTGAG 4320  
 |||||  
 Db 4380 CATGGAGCCAGCCTTCTAACGCTACCTTCCATCATTAAGTACTCTTCTGCCCTTGAG 4439  
 |||||  
 QY 4321 GACCTGCGAAATCCAGAACCAAGCATATGAAAAAGCGATTTAACTTCACAGAAAT 4380  
 |||||  
 Db 4440 GACCTGCGAAATCCAGAACCAAGCATATGAAAAAGCGATTTAACTTCACAGAAAT 4499  
 |||||  
 QY 4391 AGTGAATACCTTAAAGCCAGAAATCCAGAGCCCTTCTGCTGACAAAGTTTGAAGTCT 4440  
 |||||  
 Db 4500 AGTGAATACCTTAAAGCCAGAAATCCAGAGCCCTTCTGCTGACAAAGTTTGAAGTCT 4559  
 |||||  
 QY 4441 GCAGTACTTCTACCAAGTAAAAATTAAGAACCAAGAGTGAAGGTCATCCCTTTTAA 4500  
 |||||  
 Db 4560 GCAGTACTTCTACCAAGTAAAAATTAAGAACCAAGAGTGAAGGTCATCCCTTTTAA 4619  
 |||||  
 QY 4501 TGCCCATGATGATGATAGTGTGTACATGCACAGTGTGCTGGGAGTCTCAGAAATGA 4560  
 |||||  
 Db 4620 TGCCCATGATGATGATAGTGTGTACATGCACAGTGTGCTGGGAGTCTTCAAGATGA 4679  
 |||||  
 QY 4561 AACTACCATCTCAAGAGAGCTCATTAAGTTGTGATGTGAGAGGACACAGCTGAA 4620  
 |||||  
 Db 4680 AACTACCATCTCAAGAGAGCTCATTAAGTTGTGATGTGAGAGGACACAGCTGAA 4739  
 |||||  
 QY 4621 GAGTCTGGGCCACAGATTTGACGAAACATTTACTTCCCAAGGACAGATCTAGAGGA 4680  
 |||||  
 Db 4740 GAGTCTGGGCCACAGATTTGACGAAACATTTACTTCCCAAGGACAGATCTAGAGGA 4799  
 |||||  
 QY 4681 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGACCTGAATCTGATCTCT 4740  
 |||||  
 Db 4800 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGACCTGAATCTGATCTCT 4859  
 |||||  
 QY 4741 GAAGACAGAGCCCAAGCTCAGCTGTGTGGCAACATACATCTTCAACCTCTCATG 4800  
 |||||  
 Db 4860 GAAGACAGAGCCCAAGCTCAGCTGTGTGGCAACATACATCTTCAACCTCTCATG 4919  
 |||||  
 QY 4801 AAAGTCCCAATGAAGTTGGCAATCTGCCCCAGAGTCCAGCTGCTCATCTACT 4860  
 |||||  
 Db 4920 AAAGTCCCAATGAAGTTGGCAATCTGCCCCAGAGTCCAGCTGCTCATCTACT 4979  
 |||||  
 QY 4861 GATACTGCTGGGTATATGCAATGGAAGAAAGTGTGACAGAGGAGGACAGATTTGACA 4920  
 |||||  
 Db 4980 GATACTGCTGGGTATATGCAATGGAAGAAAGTGTGACAGAGGAGGACAGATTTGACA 5039  
 |||||  
 QY 4921 GCTTCAACAGAAAGGGTCAACAAAAGATGTCCATGTGTGTGCTGACCCAGAA 4980  
 |||||  
 Db 5040 GCTTCAACAGAAAGGGTCAACAAAAGATGTCCATGTGTGTGCTGACCCAGAA 5099  
 |||||  
 QY 4981 GAATTTATGCTGCTGATCAAGTTTGGCAAGAAACCAATCATTAACTTAATTAAT 5040  
 |||||  
 Db 5100 GAATTTATGCTGCTGATCAAGTTTGGCAAGAAACCAATCATTAACTTAATTAAT 5159  
 |||||  
 QY 5041 ACTGAAGAGTACTACTATGTTGTTATGAAGACAGATGTGAGTTGTGTGAAGGACA 5100  
 |||||  
 Db 5160 ACTGAAGAGTACTACTATGTTGTTATGAAGACAGATGTGAGTTGTGTGAAGGACA 5219  
 |||||  
 QY 5161 CAGCTCTATTAAGAAAGAAATGCTGATAGCATGATTTTGAAGTCAAGAGAGATGTG 5220  
 |||||  
 Db 5280 CAGCTCTATTAAGAAAGAAATGCTGATAGCATGATTTTGAAGTCAAGAGAGATGTG 5339  
 |||||  
 QY 5221 GTCAATGGAAGAAACCAAGAGTCCAAAGGAGAGAGAGATCCAGAGAGAAAGATC 5280  
 |||||  
 Db 5340 GTCAATGGAAGAAACCAAGAGTCCAAAGGAGAGAGAGATCCAGAGAGAAAGATC 5399  
 |||||  
 QY 5381 TTCAGGGGGCTAGAAATCTGTGTATGGCCCTTCAACCAATGCCCCACAGATCAACTG 5340

Db 5400 TTCAGGGGGCTAGAAATCTGTGTATGGGCCCTTCAACCAATGCCCCACAGATCAACTG 5459  
 |||||  
 QY 5341 GAATGAGTGTACAGCTGTGTGTCTCTGTGTGAAGAGCTTTATCATTCACCTT 5400  
 |||||  
 Db 5460 GAATGAGTGTACAGCTGTGTGTCTCTGTGTGAAGAGCTTTATCATTCACCTT 5519  
 |||||  
 QY 5401 GGCACAGGTGTCCACCAATGTGTGTGACGACAGATGTGCTGACAGAGCAATGGC 5460  
 |||||  
 Db 5520 GGCACAGGTGTCCACCAATGTGTGTGACGACAGATGTGCTGACAGAGCAATGGC 5579  
 |||||  
 QY 5461 TTTCATGCAATTTGGGCAAGTGTGTGAGGACACCTGTGTGACCCGAGAGTGGTTGAC 5520  
 |||||  
 Db 5580 TTTCATGCAATTTGGGCAAGTGTGTGAGGACACCTGTGTGACCCGAGAGTGGTTGAC 5639  
 |||||  
 QY 5521 AGTGTAGCACTTACCAGTCCAGAGCTGACACCTACTGATACCCAGATCCCCAC 5580  
 |||||  
 Db 5640 AGTGTAGCACTTACCAGTCCAGAGCTGACACCTACTGATACCCAGATCCCCAC 5699  
 |||||  
 QY 5581 AGCCACTAC 5589  
 |||||  
 Db 5700 AGCCACTAC 5708

RESULT 14  
 LOCUS 181034 5914 bp DNA linear PAT 10-JUN-1998  
 DEFINITION Sequence 1 from patent US 5710001.  
 ACCESSION 181034  
 VERSION 181034.1 GI:3209324  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 5914)  
 Skolnick,M.H., Goldgar,D.E., Miki,Y., Swenson,J., Kamb,A.,  
 Harshman,K.D., Shattuck-Eidens,D.M., Tavtigian,S.V., Wiseman,R.W.,  
 and Futreal,P.Andrew.  
 17q-Linked breast and ovarian cancer susceptibility gene  
 Patent: US 5710001-A 1 20-JAN-1998;  
 FEATURES  
 source Location/Qualifiers  
 1..5914  
 /organism="unknown"  
 BASE COUNT. 2006 a 1156 c 1316 g 1436 t  
 ORIGIN  
 Query Match 99.9%; Score 5585.8; DB 6; Length 5914;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 5587; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATGGAATTAATCTGCTCGGCTTGAGAGAGTACAAAATGCAATTAATGCTATGCAAGAA 60  
 |||||  
 Db 120 ATGGAATTAATCTGCTCGGCTTGAGAGAGTACAAAATGCTATTAATGCTATGCAAGAA 179  
 |||||  
 QY 61 ATCTTAGAGTGTCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 120  
 |||||  
 Db 180 ATCTTAGAGTGTCCATCTGCTGAGTGTGATCAAGAACCTGTCTCCACAAGTGTGAC 239  
 |||||  
 QY 121 CACATATTTGCCAATTTTGCATGCTGAACCTCTCAACGAGAAAGGGCCTTCACAG 180  
 |||||  
 Db 240 CACATATTTGCCAATTTTGCATGCTGAACCTCTCAACGAGAAAGGGCCTTCACAG 299  
 |||||  
 QY 181 TGTCTTATATTAAGATGATATACCAAAAAGAGCTACAAAGAAAGTACGATTTAGT 240  
 |||||  
 Db 300 TGTCTTATATTAAGATGATATACCAAAAAGAGCTACAAAGAAAGTACGATTTAGT 359  
 |||||  
 QY 241 CAACCTTGTGAAGAGCTATTGAAATATCAATTTGTGCTTTAGCTTGACAGGTTTGGAG 300  
 |||||  
 Db 360 CAACCTTGTGAAGAGCTATTGAAATATCAATTTGTGCTTTTACGCTTGACAGGTTTGGAG 419  
 |||||  
 QY 301 TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAATTAACCTCTCTGAAACATCTAAAGAT 360  
 |||||  
 Db 420 TATGCAAAACAGCTATTAATTTTGCAAAAAAGAAATTAACCTCTCTGAAACATCTAAAGAT 479



QY	361	GAAGTTCTATCATCCAAAGATGATGGGCTACAGAAACCGTGCCAAAGACCTTCTACAGAGT	420
Db	480	GAAGTTCTATCATCCAAAGATGATGGGCTACAGAAACCGTGCCAAAGACCTTCTACAGAGT	539
QY	421	GAACCCGAAATTCCTTCCCTGCAGGAAACCGCTCAGTGTGCCAAGCTCTACACTTGGA	480
Db	540	GAACCCGAAATTCCTTCCCTGCAGGAAACCGCTCAGTGTGCCAAGCTCTACACTTGGA	599
QY	481	ACTGTGAACTCTGAGAGCAAAAGACGGGTATACAACTCAAAAGACCTGTGTACATT	540
Db	600	ACTGTGAACTCTGAGAGCAAAAGACGGGTATACAACTCAAAAGACCTGTGTACATT	659
QY	541	GAATTGGATGTGATTTCTTCTGGAAGATACCGTTAATTAAGGCACTATTGCAAGTGGGA	600
Db	660	GAATTGGATGTGATTTCTTCTGGAAGATACCGTTAATTAAGGCACTATTGCAAGTGGGA	719
QY	601	GATCAAGAAATGTATACAAATACCCCTCAAGAAACAGAGATGAATACAGTTGGATTC	660
Db	720	GATCAAGAAATGTATACAAATACCCCTCAAGAAACAGAGATGAATACAGTTGGATTC	779
QY	661	GCAAAAAAGGCTGCTTGTGAATTTTCTGAGACGGTGTACAAATACTGAAATCATCA	720
Db	780	GCAAAAAAGGCTGCTTGTGAATTTTCTGAGACGGTGTACAAATACTGAAATCATCA	839
QY	721	CCCAATTAATTAATTTGATGACACCACTAGAAAGGTGAGCTGAGAGCATCCAGAAAG	780
Db	840	CCCAATTAATTAATTTGATGACACCACTAGAAAGGTGAGCTGAGAGCATCCAGAAAG	899
QY	781	TATCAGGGTAGTCTGTTTCAAACTTGCATGTGAGCCATGTGGCACAAATACTCATGCC	840
Db	900	TATCAGGGTAGTCTGTTTCAAACTTGCATGTGAGCCATGTGGCACAAATACTCATGCC	959
QY	841	AGCTCATTTACAGCANTGAGAAACAGAGTTTATTACTCACTAAAGACAGATGAATGTAGAA	900
Db	960	AGCTCATTTACAGCANTGAGAAACAGAGTTTATTACTCACTAAAGACAGATGAATGTAGAA	1019
QY	901	AAGGCTGAATTCCTTAATTAAGAAAGCAACAGCTGGCTAGCAAGAGCAATCAATACAGA	960
Db	1020	AAGGCTGAATTCCTTAATTAAGAAAGCAACAGCTGGCTAGCAAGAGCAATCAATACAGA	1079
QY	961	TGGCGTGAATTAAGAAACATGTATATGATATGCGGAGCTCCAGACAGAAAAAGGTA	1020
Db	1080	TGGCGTGAATTAAGAAACATGTATATGATATGCGGAGCTCCAGACAGAAAAAGGTA	1139
QY	1021	GATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGAGATTAAGCAGAACTGCATGC	1080
Db	1140	GATCTGAATGCTGATCCCTGCTGTGAGAGAAAGAAATGAGATTAAGCAGAACTGCATGC	1199
QY	1081	TCAGAGAAATCTAGAGATACTGAAAGATGTTCTTGGATTAACACTAAATAGCAGCATTCAG	1140
Db	1200	TCAGAGAAATCTAGAGATACTGAAAGATGTTCTTGGATTAACACTAAATAGCAGCATTCAG	1259
QY	1141	AAAGTTAATGAGTGGTTTTCACAGAAAGATGATACTGTAGGTTGTGAGACTCATCATGAT	1200
Db	1260	AAAGTTAATGAGTGGTTTTCACAGAAAGATGATACTGTAGGTTGTGAGACTCATCATGAT	1319
QY	1201	GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATATTTGACGCTTCTTAATAGAGGTAGAT	1260
Db	1320	GGGAGTCTGAATCAAAATGCCAAAGTAGCTGATATTTGACGCTTCTTAATAGAGGTAGAT	1379
QY	1261	GAATATTTCTGTTCTTTCAGAGAAATATGACTTACTGGCCAGTGAATCCTCATAGAGCTTTTA	1320
Db	1380	GAATATTTCTGTTCTTTCAGAGAAATATGACTTACTGGCCAGTGAATCCTCATAGAGCTTTTA	1439
QY	1321	ATATGTAAATATGAAAGAGTCACTCCCAATCAGTAGAGATTAATTTGAAGACAAATA	1380
Db	1440	ATATGTAAATATGAAAGAGTCACTCCCAATCAGTAGAGATTAATTTGAAGACAAATA	1499
QY	1381	TTTGGGAAAACCTATCGGAGAGAGCAAGCTCTCCCAACTTAAAGCCATGTACTGAAAT	1440
Db	1500	TTTGGGAAAACCTATCGGAGAGAGCAAGCTCTCCCAACTTAAAGCCATGTACTGAAAT	1559

QY	1441	CTAATTATAGAGACATTGTTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAAT	1500
Db	1560	CTAATTATAGGAGCGATTGTTACTGAGCCACAGATTAATACAAGAGCGTCCCTCACAAAT	1619
QY	1501	AAATTAAAGCGTAAAGAGACCTTACATCAGAGCCCTCATCTCGAGGATTTTATCAAGAA	1566
Db	1620	AAATTAAAGCGTAAAGAGAGACCTTACATCAGAGCCCTCATCTCGAGGATTTTATCAAGAA	1679
QY	1561	GCAGATTGGGAGTTCCAAAGACCTCTTAATGATTAATTCAGGGGAACTAACCAACGGAG	1620
Db	1680	GCAGATTGGGAGTTCCAAAGACCTCTTAATGATTAATTCAGGGGAACTAACCAACGGAG	1739
QY	1621	CAGAAATGTCAGATGATTAATTAATTAATAGTGGTCATGAGAAATTAACAAAGGTGAT	1680
Db	1740	CAGAAATGTCAGATGATTAATTAATTAATAGTGGTCATGAGAAATTAACAAAGGTGAT	1799
QY	1681	TCTATTGAGATGAGAAAAATCCTTAACCAATAGATTAATCTCGAAAAAGAAATCTGCTTC	1740
Db	1800	TCTATTGAGATGAGAAAAATCCTTAACCAATAGATTAATCTCGAAAAAGAAATCTGCTTC	1859
QY	1741	AAAAAGAGTGAACCTATTAAGCAGCGATTAAGCAATTAATGGAATCTGGAATTTAAATATC	1800
Db	1860	AAAAAGAGTGAACCTATTAAGCAGCGATTAAGCAATTAATGGAATCTGGAATTTAAATATC	1919
QY	1801	CACAATTCAAAAAGCAGCTTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	1860
Db	1920	CACAATTCAAAAAGCAGCTTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	1979
QY	1861	CATGGCGCTTGAACCTATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	1920
Db	1980	CATGGCGCTTGAACCTATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2039
QY	1921	ATTGATTAATGTTCTGAGCAGTGAAGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	1980
Db	2040	ATTGATTAATGTTCTGAGCAGTGAAGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2099
QY	1981	AGGCACAGCAGAAACCTTAACCTATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2040
Db	2100	AGGCACAGCAGAAACCTTAACCTATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2159
QY	2041	AGTAACAAGCCAAATGAACAGACAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2100
Db	2160	AGTAACAAGCCAAATGAACAGACAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2219
QY	2101	AAGTTAAACAANTGACCTGCTGTTCTTCTAAGTGTCAATTAAGCAATTAAGCAATTAAGCAAT	2160
Db	2220	AAGTTAAACAANTGACCTGCTGTTCTTCTAAGTGTCAATTAAGCAATTAAGCAATTAAGCAAT	2279
QY	2161	TTTGTCAATCCTAGGCTTCCAAAGAGAAACCAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2220
Db	2280	TTTGTCAATCCTAGGCTTCCAAAGAGAAACCAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2339
QY	2221	TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAACT	2280
Db	2340	TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTCCAAACT	2399
QY	2281	GAAAGATCTGTAAGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2340
Db	2400	GAAAGATCTGTAAGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2459
QY	2341	GAAAGATCTGTTACTGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2400
Db	2460	GAAAGATCTGTTACTGAGATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT	2519
QY	2401	TGTGTGATCAGTGTGACATTTTGAACCCCAAGGACTAATTCATGTTGTTCCAAA	2460
Db	2520	TGTGTGATCAGTGTGACATTTTGAACCCCAAGGACTAATTCATGTTGTTCCAAA	2579
QY	2461	GATATATGAATTAAGACAGAGAGGCTTTTAAGTATCATTTTGGACATTAAGCAATTAAGCAAT	2520
Db	2580	GATATATGAATTAAGACAGAGAGGCTTTTAAGTATCATTTTGGACATTAAGCAATTAAGCAAT	2639
QY	2521	CGGGAACCAACATTAAGAAATGAAGAAAGTAACTTGAATGCTGATTTTTCAGAAATTA	2580

Db	2640	CGGGAACACACATRGAATGGAAGAAAGTAACTGATGCTCAGTATTTGCAGATRCA	2699
QY	2581	TTCAAGGTTTCAAAGCGCAGCTCATTTTGCTGTGTTTCAATTCAGAAATGCAGAGAG	2640
Db	2700	TTCAAGGTTTCAAAGCGCAGCTCATTTTGCTGTGTTTCAATTCAGAAATGCAGAGAG	2759
QY	2641	GAATGTGACAACTTCTCTGCCACTCTGTGGTCTTAAAGAAACAAGTCCAAAAGTCACT	2700
Db	2760	GAATGTGACAACTTCTCTGCCACTCTGTGGTCTTAAAGAAACAAGTCCAAAAGTCACT	2819
QY	2701	TTTGAAATGTGAACAAAAGGAAGAAATCAAGAAAGAAATGAGTCAATATCAAGCTGTA	2760
Db	2820	TTTGAAATGTGAACAAAAGGAAGAAATCAAGAAAGAAATGAGTCAATATCAAGCTGTA	2879
QY	2761	CAGACAGTTAATATCACTCAGAGGCTTTCCTGTGTTGTCGAAAAGATAAAGCCAGTTGAT	2820
Db	2880	CAGACAGTTAATATCACTCAGAGGCTTTCCTGTGTTGTCGAAAAGATAAAGCCAGTTGAT	2939
QY	2821	AATGCGCAATATAGTATCAAAAGAGAGGCTTAGTGTGTCTATATCACTCACTTAGAGGC	2880
Db	2940	AATGCGCAATATAGTATCAAAAGAGAGGCTTAGTGTGTCTATATCACTCACTTAGAGGC	2999
QY	2881	AACGAACTGAGTCATCTACTCTCAAAATPAACATGAGACTTTTACAAAACCCATTCGATA	2940
Db	3000	AACGAACTGAGTCATCTACTCTCAAAATPAACATGAGACTTTTACAAAACCCATTCGATA	3059
QY	2941	CCACCACTTTTCCCATCAAGTCACTTTGTTAAACTAAATGTAAGAAAATCTGCTAGAG	3000
Db	3060	CCACCACTTTTCCCATCAAGTCACTTTGTTAAACTAAATGTAAGAAAATCTGCTAGAG	3119
QY	3001	GAAACCTTTGAGGAACATTCATATGTCACCTAAAGAGAAATGGAAATGAGACATTTCCA	3060
Db	3120	GAAACCTTTGAGGAACATTCATATGTCACCTAAAGAGAAATGGAAATGAGACATTTCCA	3179
QY	3061	AGTACAGTGAACACAATTAGCCGTAATPAACATTTAGAGAAATGTTTTAAAGAGCCAGC	3120
Db	3180	AGTACAGTGAACACAATTAGCCGTAATPAACATTTAGAGAAATGTTTTAAAGAGCCAGC	3239
QY	3121	TCAAGCAATATTAATGAAGTGGTTCAGTCACTAATGAAGTGGGCTCCAGTATTATGAA	3180
Db	3240	TCAAGCAATATTAATGAAGTGGTTCAGTCACTAATGAAGTGGGCTCCAGTATTATGAA	3299
QY	3181	ATAGGTTCCATGATGAAAAACATTCACAGCAACTAGTAGAAAACAGAGGCCAAATTG	3240
Db	3300	ATAGGTTCCATGATGAAAAACATTCACAGCAACTAGTAGAAAACAGAGGCCAAATTG	3359
QY	3241	AATGCTATGCTTAATTAAGAGGCTTTGCAACCTGAGGCTTAATTAACAAGCTTCCAGGA	3300
Db	3360	AATGCTATGCTTAATTAAGAGGCTTTGCAACCTGAGGCTTAATTAACAAGCTTCCAGGA	3419
QY	3301	AGTAATTGTAAAGCTCCTGAAATAAAAAAGCAGAAATATGAAGAAATAGTTCAAGCTGT	3360
Db	3420	AGTAATTGTAAAGCTCCTGAAATAAAAAAGCAGAAATATGAAGAAATAGTTCAAGCTGT	3479
QY	3361	AATACAGATTTCTCTCCATATCTGATTTAGATTAACATTAGAACGCTATGGGAATAGT	3420
Db	3480	AATACAGATTTCTCTCCATATCTGATTTAGATTAACATTAGAACGCTATGGGAATAGT	3539
QY	3421	CATACATCTCAGGTTGTGTTGTAACACGCTGATACCGTTTAGATGATGGTGAATTAAG	3480
Db	3540	CATACATCTCAGGTTGTGTTGTAACACGCTGATACCGTTTAGATGATGGTGAATTAAG	3599
QY	3481	GAAGATACTAGTTTGTGCGAAATGACATTAAGGAAAGTTGCTGTTTTAGCAAAAGC	3540
Db	3600	GAAGATACTAGTTTGTGCGAAATGACATTAAGGAAAGTTGCTGTTTTAGCAAAAGC	3659
QY	3541	GTCAGAAAGAGAGAGCTTAGCAGGAGTCTAGCCCTTACACCAATACACATTTGGCTAG	3600
Db	3660	GTCAGAAAGAGAGAGCTTAGCAGGAGTCTAGCCCTTACACCAATACACATTTGGCTAG	3719
QY	3601	GGTTAACCGAAGAGGGCCAGAAATTAAGTCTCAGAGAGAACTTATCTAGTAGGAT	3660

Db	3720	GGTTACCGAAGGGGCCAAGAATTAAGTCTCAGAAAGAACTTATCTAGTAGAGT	3779
Qy	3661	GAAGAGCTCCCTGCTCCCAACACTGTTATTGGTAAAGTAACAAATATACCTTCTCAG	3720
Db	3780	GAGAGAGCTTCCCTGCTCCCAACACTGTTATTGGTAAAGTAACAAATATACCTTCTCAG	3835
Qy	3721	TCTACTAGGCATAGACCGCTTGCTACCGAGTCTGCTGAAGAACACAGAGAGAAATTTA	3780
Db	3840	TCTACTAGGCATAGACCGCTTGCTACCGAGTCTGCTGAAGAACACAGAGAGAAATTTA	3899
Qy	3781	TTATCATTTGAGAATAGCTTTAAATAGCTAGCACTAACCGAGTAATATTTGGCAAGGACATCT	3840
Db	3900	TTATCATTTGAGAATAGCTTTAAATAGCTAGCACTAACCGAGTAATATTTGGCAAGGACATCT	3955
Qy	3841	CAGGAACATCACCTTAGTAGAGAAACAATGTTCTGCTAGCTGTTTCTTCACAGTGC	3900
Db	3960	CAGGAACATCACCTTAGTAGAGAAACAATGTTCTGCTAGCTGTTTCTTCACAGTGC	4019
Qy	3901	AGTGAATTTGGAAGACTTGTACCTCCAAATACCAACACCCAGAGATCCCTTCTGATTTGGTCTCT	3966
Db	4020	AGTGAATTTGGAAGACTTGTACCTCCAAATACCAACACCCAGAGATCCCTTCTGATTTGGTCTCT	4079
Qy	3961	TCCAAACAATAGAGCATCAGTCTGTAAGACCAGGAGATTGGTCTGATGACAAAGAAATTG	4020
Db	4080	TCCAAACAATAGAGCATCAGTCTGTAAGACCAGGAGATTGGTCTGATGACAAAGAAATTG	4139
Qy	4021	GTTTCAGATGATGAAGAAAGAGAACGGGCTTGGAGAATAATATCAAGAAGACAAAGC	4080
Db	4140	GTTTCAGATGATGAAGAAAGAGAACGGGCTTGGAGAATAATATCAAGAAGACAAAGC	4199
Qy	4081	ATGGAATTAACACTTAGTGAAGACACATCTGGGTGTGAGAGTGAACAACAGGCTCTCTAA	4140
Db	4200	ATGGAATTAACACTTAGTGAAGACACATCTGGGTGTGAGAGTGAACAACAGGCTCTCTAA	4259
Qy	4141	GACTGCTCAGGGCTATCCCTCTCAGAGTACATTTTAACCACTCAGCAAGGATACATG	4200
Db	4260	GACTGCTCAGGGCTATCCCTCTCAGAGTACATTTTAACCACTCAGCAAGGATACATG	4319
Qy	4201	CAACATTAACCTGATTAAGCTCCAGCAGAAATGCGTGAATGAAGCTGTGTAGAACAG	4260
Db	4320	CAACATTAACCTGATTAAGCTCCAGCAGAAATGCGTGAATGAAGCTGTGTAGAACAG	4379
Qy	4261	CATGGAGACCGCCTTCAACAGCTACCCCTTCATCAATAGGAGTCTCTTGCCCTTGAG	4320
Db	4380	CATGGAGACCGCCTTCAACAGCTACCCCTTCATCAATAGGAGTCTCTTGCCCTTGAG	4439
Qy	4321	GACCTGCGAAATCCGAAACAAAGCCATCAGAAAGAGATTAATTAATTCACAGAAAGT	4380
Db	4440	GACCTGCGAAATCCGAAACAAAGCCATCAGAAAGAGATTAATTAATTCACAGAAAGT	4499
Qy	4381	AGTGAATACCTTATAGCCAGAAATCCAGAAAGGCTTTCGTGACAACTTGTAGGTGCT	4440
Db	4500	AGTGAATACCTTATAGCCAGAAATCCAGAAAGGCTTTCGTGACAACTTGTAGGTGCT	4559
Qy	4441	GCAAGTATGTTCTACCAAGTAAAAATTAABAACCAAGAGTGGAAAGGTCAATCCCTTCTTAA	4500
Db	4560	GCAAGTATGTTCTACCAAGTAAAAATTAABAACCAAGAGTGGAAAGGTCAATCCCTTCTTAA	4619
Qy	4501	TGCCCATATTAAGATAGGAGGGTACATGCAAGTTCTCTGGGAGTCTTCAGAAATGA	4560
Db	4620	TGCCCATATTAAGATAGGAGGGTACATGCAAGTTCTCTGGGAGTCTTCAGAAATGA	4679
Qy	4561	AACCTAACCATTCAGAGAGACTCATTAAGTTGTTGATGTGAGAGCAACACACTGGAA	4620
Db	4680	AACCTAACCATTCAGAGAGACTCATTAAGTTGTTGATGTGAGAGCAACACACTGGAA	4739
Qy	4621	GAGTCTGGGCACACGATTTGACGGAAACATCTTACTTGGCCAAAGCAAGATCTAGAGGA	4680
Db	4740	GAGTCTGGGCACACGATTTGACGGAAACATCTTACTTGGCCAAAGCAAGATCTAGAGGA	4799
Qy	4661	ACCCCTTAACCTGGAATCTGGAATACAGCTCTTCTCTATGACCTGGAATCTGATCTTCT	4740
Db	4800	ACCCCTTAACCTGGAATCTGGAATACAGCTCTTCTCTATGACCTGGAATCTGATCTTCT	4859



```

QY 4741 GAAGACAGAGCCCGAGAGTCTGTTGGCAACATCCATCTTCAACTCTGCATTG 4800
    |||||
Db 4860 GAAGACAGAGCCCGAGAGTCTGTTGGCAACATCCATCTTCAACTCTGCATTG 4919
    |||||
QY 4801 AAGATTCCCAATTTGAAGTTCGAGAACTGCGCCAGAGCCAGCTGCTCTCATACT 4860
    |||||
Db 4920 AAGATTCCCAATTTGAAGTTCGAGAACTGCGCCAGAGCTGCTCTCATACT 4979
    |||||
QY 4861 GATACGTCTGGTATATGCAATGAGAAGAGTGTAGAGAGGAGAGAGCAATGACA 4920
    |||||
Db 4980 GATACGTCTGGTATATGCAATGAGAAGAGTGTAGAGAGGAGAGAGCAATGACA 5039
    |||||
QY 4921 GCTTCAACAGAAAGGCTCAACAAAGATGTCATGTTGCTGCTGACCCAGAA 4980
    |||||
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGATGTCATGTTGCTGCTGACCCAGAA 5099
    |||||
QY 4981 GAATTATGCTGTGACAAAGTTGGCAACAAACACCATCACTTAATCTAAT 5040
    |||||
Db 5100 GAATTATGCTGTGACAAAGTTGGCAACAAACACCATCACTTAATCTAAT 5159
    |||||
QY 5041 ACTGAGAGACTACTGATGTTGATGAAGAGATGATGTTGTTGTGAGAGGACA 5100
    |||||
Db 5160 ACTGAGAGACTACTGATGTTGATGAAGAGATGATGTTGTTGTGAGAGGACA 5219
    |||||
QY 5101 CTGAATATTTCTAGAAATGCGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGACC 5160
    |||||
Db 5220 CTGAATATTTCTAGAAATGCGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGACC 5279
    |||||
QY 5161 CAGTCTATTTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGTG 5220
    |||||
Db 5280 CAGTCTATTTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGTG 5339
    |||||
QY 5221 GTCAATGAGAAGAACCCAGAGTCCAAAGCGAGAGAAATCCAGAGACAGAAAGATC 5280
    |||||
Db 5340 GTCAATGAGAAGAACCCAGAGTCCAAAGCGAGAGAAATCCAGAGACAGAAAGATC 5399
    |||||
QY 5281 TTCAGGGGGCTAGAAATCTGTTGCTATGGGCTTCCACCAAGATGCCACAGATCACTG 5340
    |||||
Db 5400 TTCAGGGGGCTAGAAATCTGTTGCTATGGGCTTCCACCAAGATGCCACAGATCACTG 5459
    |||||
QY 5341 GAATGATGTGTACAGTGTGTGTGCTTCTGTGTGTAAGAGCTTTTCAATCACTTACCTT 5400
    |||||
Db 5460 GAATGATGTGTACAGTGTGTGTGCTTCTGTGTGTAAGAGCTTTTCAATCACTTACCTT 5519
    |||||
QY 5401 GGCACAGGTGTCCACCAATTTGTTGTTGTGACGCCAGATGCCGAGACAGAGCAATGGC 5460
    |||||
Db 5520 GGCACAGGTGTCCACCAATTTGTTGTTGTGACGCCAGATGCCGAGACAGAGCAATGGC 5579
    |||||
QY 5461 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGGTGTTGGAC 5520
    |||||
Db 5580 TTCATGCAATTTGGGAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGGTGTTGGAC 5639
    |||||
QY 5521 AATGTAGCACTTACCAAGTGTGAGAGAGCTGTGACACTTACCTGATACCCAGATCCCCAC 5580
    |||||
Db 5640 AATGTAGCACTTACCAAGTGTGAGAGAGCTGTGACACTTACCTGATACCCAGATCCCCAC 5699
    |||||
QY 5581 AGCCACTAC 5589
    |||||
Db 5700 AGCCACTAC 5708
    |||||

```

```

RESULT 15
AR048660 5711 bp DNA linear PAT 29-SEP-1999
LOCUS AR048660
DEFINITION Sequence 4 from patent US 5821328.
ACCESSION AR048660
VERSION AR048660.1 GI:5971003
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 5711)

```

```

AUTHORS King, M.-C., Friedman, L., Ostermeyer, B., Rowell, S., Lynch, E.,
Szabo, C., and Lee, M.
TITLE Genetic markers for breast, ovarian, and prostatic cancer
JOURNAL Patent: US 5821328-A 4 13-OCT-1998;
FEATURES Location/Qualifiers
source 1..5711
BASE COUNT 1956 a 1099 c 1275 g 1381 t
ORIGIN
Query Match 99.9%; Score 5584.2; DB 6; Length 5711;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5586; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 ATGATTTATCTGCTCTGCGTTGAAGATGCAAAATGTCAATATGCTATGTCAGAAA 60
120 ATGATTTATCTGCTCTTGGCGTTGAAGATGCAAAATGTCAATATGCTATGTCAGAAA 179
Db 180 ATCTAGAGTGTCCATCTGTCTGAGTTGATCAAGAACCTGTCTCCAAAGGTGTAC 239
QY 61 ATCTAGAGTGTCCATCTGTCTGAGTTGATCAAGAACCTGTCTCCAAAGGTGTAC 120
    |||||
Db 180 ATCTAGAGTGTCCATCTGTCTGAGTTGATCAAGAACCTGTCTCCAAAGGTGTAC 239
    |||||
QY 121 CACATATTTTGCAAATTTTGCATGCTGAACCTTCTACCCAGAGAAAGGCTTACAG 180
    |||||
Db 240 CACATATTTTGCAAATTTTGCATGCTGAACCTTCTACCCAGAGAAAGGCTTACAG 299
    |||||
QY 181 TGTCTTTATGTAGATATATTAACCAAGAGGCTTCAAGAAATGTGAGATTTAGT 240
    |||||
Db 300 GTCTCTTTATGTAGATATATTAACCAAGAGGCTTCAAGAAATGTGAGATTTAGT 359
    |||||
QY 241 CACATTTTGAAGAGCTATTTGAAATCATTTGTGCTTTTACAGTTGACACAGTTTGGAG 300
    |||||
Db 360 CACATTTTGAAGAGCTATTTGAAATCATTTGTGCTTTTACAGTTGACACAGTTTGGAG 419
    |||||
QY 301 TATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCTCTTAACATCTAAAGAT 360
    |||||
Db 420 TATGCAACAGCTATATTTTGCAGAAAGAAATTAATCTCTCTTAACATCTAAAGAT 479
    |||||
QY 361 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTCCAAAGATCTTACAGAGT 420
    |||||
Db 480 GAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTCCAAAGATCTTACAGAGT 539
    |||||
QY 421 GAACCCGAAATCTCTCTTGCAGAGAACAGTCTAGTGTCCAACTCTTAACCTTGA 480
    |||||
Db 540 GAACCCGAAATCTCTCTTGCAGAGAACAGTCTAGTGTCCAACTCTTAACCTTGA 599
    |||||
QY 481 ACTGTAGAACTCTGTGAGCAAGAGCGGATTAACCTCAAAAGACGTCTGTCAATT 540
    |||||
Db 600 ACTGTAGAACTCTGTGAGCAAGAGCGGATTAACCTCAAAAGACGTCTGTCAATT 659
    |||||
QY 541 GAATTGGATCTGATCTCTGTAAGTACCGTTAATGAAGCAACTAATGCAAGTGGGA 600
    |||||
Db 660 GAATTGGATCTGATCTCTGTAAGTACCGTTAATGAAGCAACTAATGCAAGTGGGA 719
    |||||
QY 601 GATCAAGATTTGTACAAATCAACCCCTCAAGAGAACCGAGATGAATTCAGTTGATTCT 660
    |||||
Db 720 GATCAAGATTTGTACAAATCAACCCCTCAAGAGAACCGAGATGAATTCAGTTGATTCT 779
    |||||
QY 661 GCAAAAAAGGCTGCTGTGTAATTTTCTGAGAGGATGTACAAATTAATCAATCATCA 720
    |||||
Db 780 GCAAAAAAGGCTGCTGTGTAATTTTCTGAGAGGATGTACAAATTAATCAATCATCA 839
    |||||
QY 721 CCCAGTAATATGATTTGAAACACCTGAGAGGCGTGTGAGAGGATCCAGAAAG 780
    |||||
Db 840 CCCAGTAATATGATTTGAAACACCTGAGAGGCGTGTGAGAGGATCCAGAAAG 899
    |||||
QY 781 TATCAGGGTAGTCTGTTTCAAACTTGTGAGAGGATGTGAGCAAAATTAATCAATGCC 840
    |||||
Db 900 TATCAGGGTAGTCTGTTTCAAACTTGTGAGAGGATGTGAGCAAAATTAATCAATGCC 959
    |||||
QY 841 AGCTCATTAACAGATGAGACAGAGCTTATTACTACTAAAGACAGATGATGTAGAA 900
    |||||
Db 960 AGCTCATTAACAGATGAGACAGAGCTTATTACTACTAAAGACAGATGATGTAGAA 1019
    |||||

```

QY 901 AAGCTGATTCCTGTAATAAAAGCAACAGCCTGGCTAGCAAGGAGCCACATATACAGA 960  
DB 1020 AAGCTGATTCCTGTAATAAAAGCAACAGCCTGGCTAGCAAGGAGCCACATATACAGA 1079  
QY 961 TGGCTGGAAGTAAAGAAACATGTAAATGATAGCGGACTCCAGCAGCAAAAAAGGTA 1020  
DB 1080 TGGCTGGAAGTAAAGAAACATGTAAATGATAGCGGACTCCAGCAGCAAAAAAGGTA 1139  
QY 1021 GATCGAATGCGATCCCTGTGTAGAGAAAAAGATGAAGAGAACTGCCATGTC 1080  
DB 1140 GATCGAATGCGATCCCTGTGTAGAGAAAAAGATGAAGAGAACTGCCATGTC 1199  
QY 1081 TCAGAGAAATCCAGAGATACAGAGATGTTCTTGATTAACACTAAATAGCAGCATTCAG 1140  
DB 1200 TCAGAGAAATCCAGAGATACAGAGATGTTCTTGATTAACACTAAATAGCAGCATTCAG 1259  
QY 1141 AAAGTAAATGAGTGGTTTTCCAGAAAGTGAATGTTAGTTCTGATGACTCAGATGAT 1200  
DB 1260 AAAGTAAATGAGTGGTTTTCCAGAAAGTGAATGTTAGTTCTGATGACTCAGATGAT 1319  
QY 1201 GGGGAGTCTGATCAAAATGCCAAAGTAGCTGATGATTTGAGAGTTCTAAATAGAGAT 1260  
DB 1320 GGGGAGTCTGATCAAAATGCCAAAGTAGCTGATGATTTGAGAGTTCTAAATAGAGAT 1379  
QY 1261 GAATATTCCTGTTCTTCAGAGAAAAATGACTTACTGCGCAGTATCCTCATGAGGCTTAA 1320  
DB 1380 GAATATTCCTGTTCTTCAGAGAAAAATGACTTACTGCGCAGTATCCTCATGAGGCTTAA 1439  
QY 1321 ATATGTAAGTGAAGAGATTCCTCCAAATAGTAGAGTAAATTTGAGACAAATA 1380  
DB 1440 ATATGTAAGTGAAGAGATTCCTCCAAATAGTAGAGTAAATTTGAGACAAATA 1499  
QY 1381 TTGAGGAAACCTATCGAGAGAGCAAGCCTCCCAACTTAAGCATGATTAATGTAAT 1440  
DB 1500 TTGAGGAAACCTATCGAGAGAGCAAGCCTCCCAACTTAAGCATGATTAATGTAAT 1559  
QY 1441 CTAAATATAGAGACATTTGTTACTGAGCCACAGATTAATCAAGAGCCTCCCTCAAAAT 1500  
DB 1560 CTAAATATAGAGACATTTGTTACTGAGCCACAGATTAATCAAGAGCCTCCCTCAAAAT 1619  
QY 1501 AAATTTAAAGCCTAAAAGAGAGACCTACATCAGGCTTCATCCTGAGAGATTTTCAAGAAA 1560  
DB 1620 AAATTTAAAGCCTAAAAGAGAGACCTACATCAGGCTTCATCCTGAGAGATTTTCAAGAAA 1679  
QY 1561 GCAGATTTGGCAGTTCAAAAGACCTCCTGAAATGATAAATCAGGGAATTAACCAACGAG 1620  
DB 1680 GCAGATTTGGCAGTTCAAAAGACCTCCTGAAATGATAAATCAGGGAATTAACCAACGAG 1739  
QY 1621 CAGAAATGCTCAAGTGAATTAATTAATAGTGTGATGAGAAATAAAACAAAGGTGAT 1680  
DB 1740 CAGAAATGCTCAAGTGAATTAATTAATAGTGTGATGAGAAATAAAACAAAGGTGAT 1799  
QY 1681 TCTATTCAAGATGAGAAAAATCCTAACCAATAGATACATCCTGAAAAAGATCTGCTTTC 1740  
DB 1800 TCTATTCAAGATGAGAAAAATCCTAACCAATAGATACATCCTGAAAAAGATCTGCTTTC 1859  
QY 1741 AAAAGGAAAGCTGAACCTATTAAGCAGAGATTAATGAATATGGAATGGAATTAATATTC 1800  
DB 1860 AAAAGGAAAGCTGAACCTATTAAGCAGAGATTAATGAATATGGAATGGAATTAATATTC 1919  
QY 1801 CACAATTCMAAAGACCTAAAAGAAATAGGCTGAGAGAGAGTCTTTCACAGGATATTC 1860  
DB 1920 CACAATTCMAAAGACCTAAAAGAAATAGGCTGAGAGAGAGTCTTTCACAGGATATTC 1979  
QY 1861 CATCGCTGTAAGTAACTAGTGAATGAATCTAAGCCCACTAATTTGATGTAATTCGAA 1920  
DB 1980 CATCGCTGTAAGTAACTAGTGAATGAATCTAAGCCCACTAATTTGATGTAATTCGAA 2039  
QY 1921 ATTGATAGTGTCTTAGCAGTGAAGAGATTAAGAAAAAGTAAACCAATGCGCAGTC 1980  
DB 2040 ATTGATAGTGTCTTAGCAGTGAAGAGATTAAGAAAAAGTAAACCAATGCGCAGTC 2099

QY 1981 AGGCACAGCAGAAACCTACATCTAGAGAGTAAGAAGTGAACCTGCACTGAGCCAGAAG 2040  
DB 2100 AGGCACAGCAGAAACCTACATCTAGAGAGTAAGAAGTGAACCTGCACTGAGCCAGAAG 2159  
QY 2041 AGTAACAGCCCAATGAACAGACAGTAAGAAAACATGACAGTATCTTTCCAGAGCTG 2100  
DB 2160 AGTAACAGCCCAATGAACAGACAGTAAGAAAACATGACAGTATCTTTCCAGAGCTG 2219  
QY 2101 AAGTTAACAAATGCAACCTGCTTCTTCTTACTAAGTGTCAATTAACAGCACTTAAGAA 2160  
DB 2220 AAGTTAACAAATGCAACCTGCTTCTTCTTACTAAGTGTCAATTAACAGCACTTAAGAA 2279  
QY 2161 TTGTCAATCTAGCCTTCCAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGTG 2220  
DB 2280 TTGTCAATCTAGCCTTCCAGAGAGAAAAAGAGAACTAGAAACAGTTAAAGTG 2339  
QY 2221 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGTTTTGCAACT 2280  
DB 2340 TCTAATTAATGCTGAAGACCCCAAGATCTCATGTTAACTGAGAAAGGTTTTGCAACT 2399  
QY 2281 GAAAGATCTGTAGAGAGTGAAGAGATTTCAATTTGACTGCTGACTGATTAAGCAGTCA 2340  
DB 2400 GAAAGATCTGTAGAGAGTGAAGAGATTTCAATTTGACTGCTGACTGATTAAGCAGTCA 2459  
QY 2341 GAAAGATCTGTAGAGAGTGAAGAGATTTCAATTTGACTGCTGACTGATTAAGCAGTCA 2400  
DB 2460 GAAAGATCTGTAGAGAGTGAAGAGATTTCAATTTGACTGCTGACTGATTAAGCAGTCA 2519  
QY 2401 TGTGTGAGTCACTGTGACAGCATTTGAAACCCCAAGGACTAATTCATGTTGTTCCAA 2460  
DB 2520 TGTGTGAGTCACTGTGACAGCATTTGAAACCCCAAGGACTAATTCATGTTGTTCCAA 2579  
QY 2461 GATTAATAGAAATGACACAGAGGCTTTAAGTATCCATTTGGAATGAAGTTAAACAAGT 2520  
DB 2580 GATTAATAGAAATGACACAGAGGCTTTAAGTATCCATTTGGAATGAAGTTAAACAAGT 2639  
QY 2521 CGGGAACAGCATGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGACAGATACA 2580  
DB 2640 CGGGAACAGCATGAAATGGAAGAAAGTGAACCTGATGCTCAGTATTTGACAGATACA 2699  
QY 2581 TTCAAGGTTTCAAAACCGCAGTCACTGCTGTTTCAAAATCCAGGAAATGCAAGAGAG 2640  
DB 2700 TTCAAGGTTTCAAAACCGCAGTCACTGCTGTTTCAAAATCCAGGAAATGCAAGAGAG 2759  
QY 2641 GAATGTCGAACCTTCTGCGCCAGCTGCTGCTGCTTAAAGAAACCAAGTCCAAAGTCACT 2700  
DB 2760 GAATGTCGAACCTTCTGCGCCAGCTGCTGCTGCTTAAAGAAACCAAGTCCAAAGTCACT 2819  
QY 2701 TTGGAATGTGAACAAAGAGAAATCAAGAAAGATGATCTAATATCAAGCCTGTA 2760  
DB 2820 TTGGAATGTGAACAAAGAGAAATCAAGAAAGATGATCTAATATCAAGCCTGTA 2879  
QY 2761 CAGACAGTTAATATCAGTCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
DB 2880 CAGACAGTTAATATCAGTCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2939  
QY 2821 AATGCAATGATGATATCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880  
DB 2940 AATGCAATGATGATATCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2999  
QY 2881 AAGCAAACTGACATCTACTCCAAATTAACATGATGATTTTACAAAACCCATATCTGTA 2940  
DB 3000 AAGCAAACTGACATCTACTCCAAATTAACATGATGATTTTACAAAACCCATATCTGTA 3059  
QY 2941 CCACCACTTTTCCCATCAAGTCAATTTGTTTAAACCTAATGTAAGAAAAATCTGCTAGAG 3000  
DB 3060 CCACCACTTTTCCCATCAAGTCAATTTGTTTAAACCTAATGTAAGAAAAATCTGCTAGAG 3119  
QY 3001 GAAACCTTTGAGAAACATTCATGCTCACTGGAAGAGAAATGGAATGGAATGGAATTCGA 3060  
DB 3120 GAAACCTTTGAGAAACATTCATGCTCACTGGAAGAGAAATGGAATGGAATGGAATTCGA 3179  
QY 3061 AGTACAGTGAACCAATTAAGCCTGTAATTAAGAAAAATGTTTTTAAAGAGCCAGC 3120

```
|||||
Db 3180 AGTACATGAGCACAATTTAGCCGTAATTAACATTAGAGAAATGTTTTAAAGAGCCAGC 3239
QY 3121 TCAGCAATATTAATGAAGTAGTTCAGTACTAATGAGTGGGCTCAGATTAATGAA 3180
Db 3240 TCAGCAATATTAATGAAGTAGTTCAGTACTAATGAGTGGGCTCAGATTAATGAA 3239
QY 3181 ATAGGTTCCAGTGTGAAACATTCAGCAGACTAGTAGAAACAGAGGCCCAAAATG 3240
Db 3300 ATAGGTTCCAGTGTGAAACATTCAGCAGACTAGTAGAAACAGAGGCCCAAAATG 3359
QY 3241 AATGCTATGCTTAATTAATGAGGTTTTGCAACCTAGGCTTAATAACAAGTCTTGGA 3300
Db 3360 AATGCTATGCTTAATTAATGAGGTTTTGCAACCTAGGCTTAATAACAAGTCTTGGA 3419
QY 3301 AGTAATGTGAAGCATCCGTAATTAATAAGCAAGATATGAAAGTAGTTCAGACTGT 3360
Db 3420 AGTAATGTGAAGCATCCGTAATTAATAAGCAAGATATGAAAGTAGTTCAGACTGT 3479
QY 3361 AATACAGATTTCTCCATATCTGATTTAGATTAAGTAAACAGCCTATGGAGTAGT 3420
Db 3480 AATACAGATTTCTCCATATCTGATTTAGATTAAGTAAACAGCCTATGGAGTAGT 3539
QY 3421 CATGCTATCTCAGGTTTTGTTGAGACACCTGATGACTGTTAGTATGATGGAATTAAG 3480
Db 3540 CATGCTATCTCAGGTTTTGTTGAGACACCTGATGACTGTTAGTATGATGGAATTAAG 3599
QY 3481 GAAGATCTAGTTTGGTGAATAATGACATTAAGAAAGTTGCTGTTTTAGCAAAAGC 3540
Db 3600 GAAGATCTAGTTTGGTGAATAATGACATTAAGAAAGTTGCTGTTTTAGCAAAAGC 3659
QY 3541 GTCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACATTTGGCTCAG 3600
Db 3660 GTCCAGAAAGAGAGCTTAGCAGAGAGTCTAGCCCTTTCACCCATACATTTGGCTCAG 3719
QY 3601 GGTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGAGAGAT 3660
Db 3720 GGTTACCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGAGAGAT 3779
QY 3661 GAAGAGCTCCCTCTCCAAACACTTGTATTTGGTAAAGTAAACATATACCTCTCAG 3720
Db 3780 GAAGAGCTCCCTCTCCAAACACTTGTATTTGGTAAAGTAAACATATACCTCTCAG 3839
QY 3721 TCTACTAGGCAATGACACCGTTGCTACGAGTGTCTGTCTAGAAACAGAGAGAAATTA 3780
Db 3840 TCTACTAGGCAATGACACCGTTGCTACGAGTGTCTGTCTAGAAACAGAGAGAAATTA 3899
QY 3781 TTATCATTTGAAGATTAATGACTGAGTACAGTAAATATTTGGCAAGGCAATCT 3840
Db 3900 TTATCATTTGAAGATTAATGACTGAGTAAACAGGTAATATTTGGCAAGGCAATCT 3959
QY 3841 CAGGACATCACTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTTCTCCAGTGC 3900
Db 3960 CAGGACATCACTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTTCTCCAGTGC 4019
QY 3901 AGTGAATTGGAAGACTGACTGCAATATACAAACCCAGAGATCTTCTTGATTTGTTCT 3960
Db 4020 AGTGAATTGGAAGACTGACTGCAATATACAAACCCAGAGATCTTCTTGATTTGTTCT 4079
QY 3961 TCCAAACAAATGAGGCACTGAGTCTGAAAGCCAGGAGTTGGTCTGAGTGAACGAATTTG 4020
Db 4080 TCCAAACAAATGAGGCACTGAGTCTGAAAGCCAGGAGTTGGTCTGAGTGAACGAATTTG 4139
QY 4021 GTTTCAGATGATGAAGAAAGAGAGGGCTTGGAGAAATAATCAAGAAAGAGCAAGC 4080
Db 4140 GTTTCAGATGATGAAGAAAGAGAGGGCTTGGAGAAATAATCAAGAAAGAGCAAGC 4199
QY 4081 ATGGAATCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAGCAAGCCTCTCGAA 4140
Db 4200 ATGGAATCAAACTTAGTGAAGCAGCATCTGGGTGTGAGAGTGAAGCAAGCCTCTCGAA 4259
QY 4141 GACTGCTCAGGCGTATCTCTCAGAGTGAATTTAAACCACTCAGCAGAGGATCCATG 4200
|||||

Db 4260 GACTGCTCAGGCGTATCTCTCAGAGTGAATTTAAACCACTCAGCAGAGGATCCATG 4319
QY 4201 CAACATAACTGATAAAGCTCCAGAGAGAAATGGCTGAATAGAAAGTGTGTAAGCAG 4260
Db 4320 CAACATAACTGATAAAGCTCCAGAGAGAAATGGCTGAATAGAAAGTGTGTAAGCAG 4379
QY 4261 CATGGAGCCAGCCTTCTAACAGCTACCTTCCATCATTAAGTGAATCTTTCGCCCTGAG 4320
Db 4380 CATGGAGCCAGCCTTCTAACAGCTACCTTCCATCATTAAGTGAATCTTTCGCCCTGAG 4439
QY 4321 GACTCGGAATTCAGAAACAAAGCAGATCAGAAAAAGCAGTATTAATCTCAGAAAAAGT 4380
Db 4440 GACTCGGAATTCAGAAACAAAGCAGATCAGAAAAAGCAGTATTAATCTCAGAAAAAGT 4499
QY 4381 AGTAATACCTTATTAAGCAGAAATCCAGAAAGCCTTCTGCTGCAAGTTGAGTGTCT 4440
Db 4500 AGTAATACCTTATTAAGCAGAAATCCAGAAAGCCTTCTGCTGCAAGTTGAGTGTCT 4559
QY 4441 GCAGATAGTTCTACAGTAAATAAAGAAACAGAGTGAAGAGTCAATCCCTTCTAAA 4500
Db 4560 GCAGATAGTTCTACAGTAAATAAAGAAACAGAGTGAAGAGTCAATCCCTTCTAAA 4619
QY 4501 TGCCCATCATTAATGATGATAGTGTGATGACATGACAGTGTCTTGAGAGTCTTCAAAATGA 4560
Db 4620 TGCCCATCATTAATGATGATAGTGTGATGACATGACAGTGTCTTGAGAGTCTTCAAAATGA 4679
QY 4561 AACTACCATCTCAAGAGAGAGCTCATTAAGTTGTGTGAGAGAGCAGACACTGGAA 4620
Db 4680 AACTACCATCTCAAGAGAGAGCTCATTAAGTTGTGTGAGAGAGCAGACACTGGAA 4739
QY 4621 GAGTCTGGGCCACACAGATTTGACGGAACATCTTACTTGGCCAAAGCAAGATCTAGAGGA 4680
Db 4740 GAGTCTGGGCCACACAGATTTGACGGAACATCTTACTTGGCCAAAGCAAGATCTAGAGGA 4799
QY 4681 ACCCCTTACCTGGAATCTGGAATGACCTCTTCTGTATGACCTGGAATCTGATCTTCT 4740
Db 4800 ACCCCTTACCTGGAATCTGGAATGACCTCTTCTGTATGACCTGGAATCTGATCTTCT 4859
QY 4741 GAAGACAGAGCCCGAGAGTGTGTTGGCAACATTAACATCTTAACCTCTGCATTTG 4800
Db 4860 GAAGACAGAGCCCGAGAGTGTGTTGGCAACATTAACATCTTAACCTCTGCATTTG 4919
QY 4801 AAAGTCCCAATTTGAAAGTTGCAAGATCTGCCCCAGAGTCCAGCTGCTCATACTACT 4860
Db 4920 AAAGTCCCAATTTGAAAGTTGCAAGATCTGCCCCAGAGTCCAGCTGCTCATACTACT 4979
QY 4861 GATACCTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGAGGAGAAACCGAATTTGACA 4920
Db 4980 GATACCTGCTGGGTATTAATGCAATGGAAGAAAGTGTGAGCAGAGGAGAAACCGAATTTGACA 5039
QY 4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGGTGGTGTGCTGGCTGACCCCGAA 4980
Db 5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGGTGGTGTGCTGGCTGACCCCGAA 5099
QY 4981 GAATTTATGCTCGTGTACAAAGTTTGCAGAAACCAACCAATCACTTAATCAATCAAT 5040
Db 5100 GAATTTATGCTCGTGTACAAAGTTTGCAGAAACCAACCAATCACTTAATCAATCAAT 5159
QY 5041 ACTGAAGAGACTACTCATGTTGTTATGAAAAACAGATGCTGAGTTGTGTGAAACGAGCA 5100
Db 5160 ACTGAAGAGACTACTCATGTTGTTATGAAAAACAGATGCTGAGTTGTGTGAAACGAGCA 5219
QY 5101 CTGAAATATTTCTAGGAATTTGGGGAGGAAATGGGTAGTACTATTTCTGGGTGACC 5160
Db 5220 CTGAAATATTTCTAGGAATTTGGGGAGGAAATGGGTAGTACTATTTCTGGGTGACC 5279
QY 5161 CAGTCTATTAAGAAAGAAATGCTGAATGAGCATTTTGAAGTCAAGAGAGATGTG 5220
Db 5280 CAGTCTATTAAGAAAGAAATGCTGAATGAGCATTTTGAAGTCAAGAGAGATGTG 5339
QY 5221 GTCATATGAAGAAACCAAGGTCCTCAAGAGCAGACAGAAATCCAGAGACAGAAAGATC 5280
Db 5340 GTCATATGAAGAAACCAAGGTCCTCAAGAGCAGACAGAAATCCAGAGACAGAAAGATC 5399
|||||
```

```
QY 5281 TTCAGGGGGCTAGAAATCTGTTGCTATGGGCCCTTCACCAACATGCCACAGATCAACTG 5340
    |||||||
Db 5400 TTCAGGGGGCTAGAAATCTGTTGCTATGGGCCCTTCACCAACATGCCACAGATCAACTG 5459
    |||||||
QY 5341 GAATGATGATACAGCTGTGTGCTTCTGTGTGAGAGAGCTTTCATCATTACCCCTT 5400
    |||||||
Db 5460 GAATGATGATACAGCTGTGTGCTTCTGTGTGAGAGAGCTTTCATCATTACCCCTT 5519
    |||||||
QY 5401 GGCACAGATGTCACCCCAATTGATGTTGTGACAGCAGATGCCAGAGAGAGCAATGGC 5460
    |||||||
Db 5520 GGCACAGATGTCACCCCAATTGATGTTGTGACAGCAGATGCCAGAGAGAGCAATGGC 5579
    |||||||
QY 5461 TTCATGCAATTGGGCAGATGTGTGAGGCACCTGTGTGACCCGAGAGTGTGTGGAC 5520
    |||||||
Db 5580 TTCATGCAATTGGGCAGATGTGTGAGGCACCTGTGTGACCCGAGAGTGTGTGGAC 5639
    |||||||
QY 5521 AGTGTAGCACTCTACCAAGTCCAGAGAGCTGGACACCTACTGATACCCAGATCCCCAC 5580
    |||||||
Db 5640 AGTGTAGCACTCTACCAAGTCCAGAGAGCTGGACACCTACTGATACCCAGATCCCCAC 5699
    |||||||
QY 5581 AGCCACTAC 5589
    |||||||
Db 5700 AGCCACTAC 5708
```

Search completed: June 27, 2003, 01:02:14  
Job time : 13935.7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 16:33:51 ; Search time 719.151 Seconds

(without alignments)  
11536.634 Million cell updates/sec

Title: US-09-734-672-3\_COPY\_120\_5708

Perfect score: 5589

Sequence: 1 ATGATTTATCTGCTCTTCG.....AGATCCCCACAGCCACTAC 5589

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues 2111440

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCRT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCRTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5587.4	100.0	5711	9	US-09-734-672-3
2	5587.4	100.0	5711	9	US-09-982-828-5
3	5587.4	99.9	5711	9	US-09-734-672-5
4	5587.6	99.9	5711	9	US-09-982-828-3
5	5587.6	99.9	5711	9	US-10-022-819-1
6	5581	99.9	5711	9	US-09-734-672-1
7	5581	99.9	5711	9	US-09-982-828-1
8	364.6	6.5	499	9	US-09-911-904-127
9	175.4	3.1	424	10	US-09-864-761-4552
10	147	2.6	147	10	US-09-864-761-21299
11	121	2.2	121	9	US-09-818-875-661
12	121	2.2	121	9	US-09-818-875-662
13	121	2.2	121	9	US-09-818-875-665
14	121	2.2	121	9	US-09-818-875-666
15	121	2.2	121	9	US-09-818-875-669
16	121	2.2	121	9	US-09-818-875-670
17	121	2.2	121	9	US-09-818-875-673
18	121	2.2	121	9	US-09-818-875-674
19	121	2.2	121	9	US-09-818-875-677

c	20	121	2.2	121	9	US-09-818-875-678	Sequence 678, App
c	21	121	2.2	121	9	US-09-818-875-681	Sequence 681, App
c	22	121	2.2	121	9	US-09-818-875-682	Sequence 682, App
c	23	121	2.2	121	9	US-09-818-875-685	Sequence 685, App
c	24	121	2.2	121	9	US-09-818-875-686	Sequence 686, App
c	25	121	2.2	121	9	US-09-818-875-689	Sequence 689, App
c	26	121	2.2	121	9	US-09-818-875-690	Sequence 690, App
c	27	121	2.2	121	9	US-09-818-875-693	Sequence 693, App
c	28	121	2.2	121	9	US-09-818-875-694	Sequence 694, App
c	29	121	2.2	121	9	US-09-818-875-697	Sequence 697, App
c	30	121	2.2	121	9	US-09-818-875-698	Sequence 698, App
c	31	121	2.2	121	9	US-09-818-875-701	Sequence 701, App
c	32	121	2.2	121	9	US-09-818-875-702	Sequence 702, App
c	33	121	2.2	121	9	US-09-818-875-705	Sequence 705, App
c	34	121	2.2	121	9	US-09-818-875-706	Sequence 706, App
c	35	121	2.2	121	9	US-09-818-875-709	Sequence 709, App
c	36	121	2.2	121	9	US-09-818-875-710	Sequence 710, App
c	37	121	2.2	121	9	US-09-818-875-713	Sequence 713, App
c	38	121	2.2	121	9	US-09-818-875-714	Sequence 714, App
c	39	121	2.2	121	9	US-09-818-875-717	Sequence 717, App
c	40	121	2.2	121	9	US-09-818-875-718	Sequence 718, App
c	41	121	2.2	121	9	US-09-818-875-721	Sequence 721, App
c	42	121	2.2	121	9	US-09-818-875-722	Sequence 722, App
c	43	121	2.2	121	9	US-09-818-875-725	Sequence 725, App
c	44	121	2.2	121	9	US-09-818-875-726	Sequence 726, App
c	45	121	2.2	121	9	US-09-818-875-729	Sequence 729, App

#### ALIGNMENTS

RESULT 1  
US-09-734-672-3  
; Sequence 3, Application US/09734672  
; Publication No. US20020183268A1  
GENERAL INFORMATION:  
APPLICANT: Murphy, Patricia D.  
Alvares, Christoppher P.  
Critz, Brenda S.  
Olson, Sheri J.  
Schelter, Denise B.  
Zeng, Bin  
TITLE OF INVENTION: Coding Sequences of the Human  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan Lewis & Bockius LLP  
STREET: 1111 Pennsylvania Ave., N.W.  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/734,672  
FILING DATE: 03-Dec-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-No. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001

```

: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 5711 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: No. US20020183268A1 Relevant
:   TOPOLOGY: Linear
:   MOLECULE TYPE: cDNA
:   ORIGINAL SOURCE:
:     ORGANISM: Homo sapiens
:     STRAIN: BRCAl
:   POSITION IN GENOME:
:     CHROMOSOME/SEGMENT: 17
:     MAP POSITION: 17q21
: SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-734-672-3

Query Match      100.0%; Score 5587.4; DB 9; Length 5711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGATATTTATCTCTCTGCGTTGAAGATCAAAATGTCATTAATGCTATGCAGAAA 60
DB 120 ATGATATTTATCTCTCTGCGTTGAAGATCAAAATGTCATTAATGCTATGCAGAAA 179
OY 61 ATCTTAGAGTGTCCATCTGTCTGAGTTGATCAAGGAACCTGTCTCCACAAAGTGAC 120
DB 180 ATCTTAGAGTGTCCATCTGTCTGAGTTGATCAAGGAACCTGTCTCCACAAAGTGAC 239
OY 121 CACATATTTTGCATTTTGCATGCTGAACTTCTCACCAGAAAGAAAGGCCCTTCACAG 180
DB 240 CACATATTTTGCATTTTGCATGCTGAACTTCTCACCAGAAAGAAAGGCCCTTCACAG 299
OY 181 TGTCTTTATGTAGATGATATTAACCAAAAGAGCCTACAGAAATAGAGATTAGT 240
DB 300 TGTCTTTATGTAGATGATATTAACCAAAAGAGCCTACAGAAATAGAGATTAGT 359
OY 241 CAACCTGTTGAAGAGCTATTTGAAATCATTGTGCTTTTCAGCTTGACACAGGTTTGAG 300
DB 360 CAACCTGTTGAAGAGCTATTTGAAATCATTGTGCTTTTCAGCTTGACACAGGTTTGAG 419
OY 301 TATGCAAAAGCATATTTTTCGAAAAAAGAAATTAACCTCTCTGAACTCTTAAAGAT 360
DB 420 TATGCAAAAGCATATTTTTCGAAAAAAGAAATTAACCTCTCTGAACTCTTAAAGAT 479
OY 361 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTSCCAAAAGACTTCTACAGAT 420
DB 480 GAAGTTTCTATCATCCAAAGTATGGCTACAGAAACCGTSCCAAAAGACTTCTACAGAT 539
OY 421 GAACCCGAAAAATCTCTCTGCGAAGAACCAAGTCTCAGTGCATCTCTTAACTTGA 480
DB 540 GAACCCGAAAAATCTCTCTGCGAAGAACCAAGTCTCAGTGCATCTCTTAACTTGA 539
OY 481 ACTGTGAGAACTGTGAGCAAAAGAGCGGATCAACCTCAAAAGAGCTGTCTACATTT 540
DB 600 ACTGTGAGAACTGTGAGCAAAAGAGCGGATCAACCTCAAAAGAGCTGTCTACATTT 659
OY 541 GAATTTGGATCTGATTTCTTGAAGATACCGTTAATTAAGCAACTTATTGACATGTTGGA 600
DB 660 GAATTTGGATCTGATTTCTTGAAGATACCGTTAATTAAGCAACTTATTGACATGTTGGA 719
OY 601 GATCAAGAATTTGTTACAATACCCCTCAAGGAACCAAGGATGAATCACTTTGATTTCT 660
DB 720 GATCAAGAATTTGTTACAATACCCCTCAAGGAACCAAGGATGAATCACTTTGATTTCT 779
OY 661 GCAAAAAAGGCTGCTGTGAATTTTCTGAGACGAGATGTAACTACTGAACTCATCAA 720
DB 780 GCAAAAAAGGCTGCTGTGAATTTTCTGAGACGAGATGTAACTACTGAACTCATCAA 839
OY 721 CCCAGTAATTAATGATTTGGAACACACACTGAGAAAGCTGAGAGGCTCCAGAAAAG 780
DB 840 CCCAGTAATTAATGATTTGGAACACACACTGAGAAAGCTGAGAGGCTCCAGAAAAG 839
OY 781 TATCAGGGTAGTTCTGTTCAAACTTGATGTGAGCCATGTGGCAAAATTAATCTCATGCC 840
```

```

DB 900 TATCAGGGTAGTTCTGTTCAAACTTGATGTGAGCCATGTGGCAAAATTAATCTCATGCC 959
OY 841 AGCTCATTTACAGATGAGAACACACACTTTATTAATCTACTAAAGACAGAAATGATATAGA 900
DB 960 AGCTCATTTACAGATGAGAACACACACTTTATTAATCTACTAAAGACAGAAATGATATAGA 1019
OY 901 AAGCTGAATTTCTGTATTAATAAGCAAAACAGCTGTGCTGTGCAAGAGCCAACTAAACGA 960
DB 1020 AAGCTGAATTTCTGTATTAATAAGCAAAACAGCTGTGCTGTGCAAGAGCCAACTAAACGA 1079
OY 961 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACTCCAGACAGAAAAAAGGTA 1020
DB 1080 TGGGCTGGAAGTAAAGAAACATGTAATGATAGCGGACTCCAGACAGAAAAAAGGTA 1139
OY 1021 GATCTGAATTCGATCCCTGTGTGAGAGAAAAAAGTAATGGAATTAAGCAAAATGTCATGC 1080
DB 1140 GATCTGAATTCGATCCCTGTGTGAGAGAAAAAAGTAATGGAATTAAGCAAAATGTCATGC 1199
OY 1081 TCAGAGAACTCTAGATATCTGAGATGTTCTTGGATTAACCTAAATAGCAGATTGAG 1140
DB 1200 TCAGAGAACTCTAGATATCTGAGATGTTCTTGGATTAACCTAAATAGCAGATTGAG 1259
OY 1141 AAGATTAAAGTGTGTTTTCAGAGATGATGAATGTTAGTTCTGATGACTCAGATGAT 1200
DB 1260 AAGATTAAAGTGTGTTTTCAGAGATGATGAATGTTAGTTCTGATGACTCAGATGAT 1319
OY 1201 GGGGAGCTGAATCAAAATGCAAAAGTAGCTGATTTGAGCTGCTAAATAGGTTAGAT 1260
DB 1320 GGGGAGCTGAATCAAAATGCAAAAGTAGCTGATTTGAGCTGCTAAATAGGTTAGAT 1379
OY 1261 GAATTAATCTGGTCTTCAGAGAAATAGACTTACTGGCCAGTATCTCATGAGCCTTTA 1320
DB 1380 GAATTAATCTGGTCTTCAGAGAAATAGACTTACTGGCCAGTATCTCATGAGCCTTTA 1439
OY 1321 ATATGTAAAGGTAAAGAGTTCACCTCAATCAGTAGAAGTAATTAAGAAACAAATA 1380
DB 1440 ATATGTAAAGGTAAAGAGTTCACCTCAATCAGTAGAAGTAATTAAGAAACAAATA 1499
OY 1381 TTTGGGAAAACTATGGAAGGAAGGAGGAGCTCCCACTTAAGCCATGTAACTGAAAT 1440
DB 1500 TTTGGGAAAACTATGGAAGGAAGGAGGAGCTCCCACTTAAGCCATGTAACTGAAAT 1559
OY 1441 CTAATTAATAGACATTTTGTACTGAGCAGACAGATTAATCAAGAGAGTCCCTCACAAT 1500
DB 1560 CTAATTAATAGACATTTTGTACTGAGCAGACAGATTAATCAAGAGAGTCCCTCACAAT 1619
OY 1501 AAATTAAGCGTAAAGGAGACCTACATCAGGCTTCATCTGAGAGATTTTATCAAGAA 1560
DB 1620 AAATTAAGCGTAAAGGAGACCTACATCAGGCTTCATCTGAGAGATTTTATCAAGAA 1679
OY 1561 GCAGATTTGGCAGTTCAAAAAGACTCCTGAAATGATTAATCAGGGAACCTAACCAAGGAG 1620
DB 1680 GCAGATTTGGCAGTTCAAAAAGACTCCTGAAATGATTAATCAGGGAACCTAACCAAGGAG 1739
OY 1621 CAGAATGTCGAAGTGAATTAATTAATTAAGTGTGATGAGAATTAACCAAAAGGTGAT 1680
DB 1740 CAGAATGTCGAAGTGAATTAATTAATTAAGTGTGATGAGAATTAACCAAAAGGTGAT 1799
OY 1681 TCTATTTCAGATGAGAAAAATCTTAACCAATAGAACTCTGGAAGAAATCTGCTTTTC 1740
DB 1800 TCTATTTCAGATGAGAAAAATCTTAACCAATAGAACTCTGGAAGAAATCTGCTTTTC 1859
OY 1741 AAAAGGAAAGCTGAACCTAATTAAGCGAGCACTATTAACCAATTAAGCACTGAAATTAATTC 1800
DB 1860 AAAAGGAAAGCTGAACCTAATTAAGCGAGCACTAATTAACCAATTAAGCACTGAAATTAATTC 1919
OY 1801 CACAATTTCAAAAGCAGCTAAAGAAATAGGCTGAGAGAGAGTCTTCTACAGGAGCATATT 1860
DB 1920 CACAATTTCAAAAGCAGCTAAAGAAATAGGCTGAGAGAGAGTCTTCTACAGGAGCATATT 1979
OY 1861 CATGCCCTTGAACCTAGTACGTAGAAATCTAAGCCACCTAATTTGATCTGAATTTGCAA 1920
```

Db 1980 CATGGCTGMACTAGTACTAGTAGAATCTAAGCCCACTAATTGTACTGATTTGCCAA 2039  
QY 1921 ATTGATAGTTGTTCTTGTAGCAGTGAAGATTAAGAAAAAATAGTACCAACCAATGCCAGTC 1980  
Db 2040 ATTGATAGTTGTTCTTGTAGCAGTGAAGATTAAGAAAAAATAGTACCAACCAATGCCAGTC 2099  
QY 1981 AGGCACAGCAGAAACCTTACACTCATGGAAGTTAAAGAACCTGCAACTGGAGCCAAAGAG 2040  
Db 2100 AGGCACAGCAGAAACCTTACACTCATGGAAGTTAAAGAACCTGCAACTGGAGCCAAAGAG 2159  
QY 2041 AGTAAACAGCCCAATTAACAGACAGACAGTAAAGACATGACAGTATACTTTCCAGAGCTG 2100  
Db 2160 AGTAAACAGCCCAATTAACAGACAGACAGTAAAGACATGACAGTATACTTTCCAGAGCTG 2219  
QY 2101 AAGTTAAACAATGACACTGGTCTTTTACTAAGTTCAATATACAGTGAACCTTAAAGAA 2160  
Db 2220 AAGTTAAACAATGACACTGGTCTTTTACTAAGTTCAATATACAGTGAACCTTAAAGAA 2279  
QY 2161 TTTGTCAATCTAGCTTCCAG 2220  
Db 2280 TTTGTCAATCTAGCTTCCAG 2339  
QY 2221 TCTAATTAATGCTGAACACCCCAAGATCTCATGTTAAGTGAAGAGGGTTTGGCAACT 2280  
Db 2340 TCTAATTAATGCTGAACACCCCAAGATCTCATGTTAAGTGAAGAGGGTTTGGCAACT 2399  
QY 2281 GAAAGATCTGTAGAGAGTACAGTATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340  
Db 2400 GAAAGATCTGTAGAGAGTACAGTATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2459  
QY 2341 GAAAGATCTGTAGAGAGTACAGTATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400  
Db 2460 GAAAGATCTGTAGAGAGTACAGTATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2519  
QY 2401 TGTGTAGTACAGTGTGACAGTATTTGAAACCCCAAGAGAGTATTTGATGCTGCTGCTG 2460  
Db 2520 TGTGTAGTACAGTGTGACAGTATTTGAAACCCCAAGAGAGTATTTGATGCTGCTGCTG 2579  
QY 2461 GATAATAGAAATGACACAGAGAGCTTAAATGATCATTTGGAGATGAAGTTAACCAAGT 2520  
Db 2580 GATAATAGAAATGACACAGAGAGCTTAAATGATCATTTGGAGATGAAGTTAACCAAGT 2639  
QY 2521 CGGGAACACAGCATAGAAATGGAAGAGTGAACCTTGATGCTCAGTATTTGGCAGAAATAC 2580  
Db 2640 CGGGAACACAGCATAGAAATGGAAGAGTGAACCTTGATGCTCAGTATTTGGCAGAAATAC 2699  
QY 2581 TTCAAGTTCGAAGGCGCAGTATTTGCTCTGTTTTCAAATCCAGAGAAATGCGAAGAG 2640  
Db 2700 TTCAAGTTCGAAGGCGCAGTATTTGCTCTGTTTTCAAATCCAGAGAAATGCGAAGAG 2759  
QY 2641 GAATGTGAACATCTCTGCGCCACTGCGGCTCTTAAAGAAACCAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGAACATCTCTGCGCCACTGCGGCTCTTAAAGAAACCAAGTCCAAAGTCACT 2819  
QY 2701 TTTGAATGTGAACAAAG 2760  
Db 2820 TTTGAATGTGAACAAAG 2879  
QY 2761 CAGACATTAATATATCTGACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
Db 2880 CAGACATTAATATATCTGACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2939  
QY 2821 AATGCCAATGTATATCAAAAGAGAGCTAGGTTTGTCTATCATCTAGTTCAAGAGC 2880  
Db 2940 AATGCCAATGTATATCAAAAGAGAGCTAGGTTTGTCTATCATCTAGTTCAAGAGC 2999  
QY 2881 AAGGAACTGAGCTCTTATCTCCAAATTAACATGACCTTTTACAAAACCATATCTGTATA 2940  
Db 3000 AAGGAACTGAGCTCTTATCTCCAAATTAACATGACCTTTTACAAAACCATATCTGTATA 3059  
QY 2941 CCACCACTTTTCCCTCAAGTATTTGTTAAACCTAATGTAAAGAAATCTGCTAGAG 3000  
Db 3060 CCACCACTTTTCCCTCAAGTATTTGTTAAACCTAATGTAAAGAAATCTGCTAGAG 3119

QY 3001 GAAACCTTTGAGAGACATTTCAATGTCACTGAAAGAGAAATGGGAAATGAGAACATTTCCA 3060  
Db 3120 GAAACCTTTGAGAGACATTTCAATGTCACTGAAAGAGAAATGGGAAATGAGAACATTTCCA 3179  
QY 3061 AGTACAGTGAGCAATTTAGCCGTAATACATTAGAGAAATGTTTTTAAAGAGCCAGC 3120  
Db 3180 AGTACAGTGAGCAATTTAGCCGTAATACATTAGAGAAATGTTTTTAAAGAGCCAGC 3239  
QY 3121 TCAGCAATATTTATGAGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAA 3180  
Db 3240 TCAGCAATATTTATGAGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAA 3299  
QY 3181 ATAGGTTCCAGTATGAAACATTTCAAGCAGACTAGTGAAGAACAGAGGCCCAATTTG 3240  
Db 3300 ATAGGTTCCAGTATGAAACATTTCAAGCAGACTAGTGAAGAACAGAGGCCCAATTTG 3359  
QY 3241 AATGCTATGCTTATGATTAAGGGGTTTTCACACCTGAGGCTATTAACAAAGCTTCTGGA 3300  
Db 3360 AATGCTATGCTTATGATTAAGGGGTTTTCACACCTGAGGCTATTAACAAAGCTTCTGGA 3419  
QY 3301 AGTAAATTTAGCATCTCTGAATTAAGAAAGCAAGATATGAAGAGTTCAGACTGTT 3360  
Db 3420 AGTAAATTTAGCATCTCTGAATTAAGAAAGCAAGATATGAAGAGTTCAGACTGTT 3479  
QY 3361 AATACAGATTTCTCTCATATCTGATTTAGATTAAGTACAGCCTATGGGAGAGT 3420  
Db 3480 AATACAGATTTCTCTCATATCTGATTTAGATTAAGTACAGCCTATGGGAGAGT 3539  
QY 3421 CATGATCTTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGAATTAAG 3480  
Db 3540 CATGATCTTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGAATTAAG 3599  
QY 3481 GAAGATACATGTTTGTCTGAAATGACATTAAGAAAGTCTGTTTACCAAAAGC 3540  
Db 3600 GAAGATACATGTTTGTCTGAAATGACATTAAGAAAGTCTGTTTACCAAAAGC 3659  
QY 3541 GTCCAGAAAGAGAGGTTTACAGAGTCTGAGGCTTTTCAACCATACACTTTGGGCTCAG 3600  
Db 3660 GTCCAGAAAGAGAGGTTTACAGAGTCTGAGGCTTTTCAACCATACACTTTGGGCTCAG 3719  
QY 3601 GGTTCAGGAGAGGGGCCAAGAAATTAAGTCTCTCAGAGAGAGAGAGAGAGAGAGAGAT 3660  
Db 3720 GGTTCAGGAGAGGGGCCAAGAAATTAAGTCTCTCAGAGAGAGAGAGAGAGAGAGAGAT 3779  
QY 3661 GAAGAGCTTCCCTGCTTCCAACTGTTATTTGGTAAAGTAAACATATACCTTCTCAG 3720  
Db 3780 GAAGAGCTTCCCTGCTTCCAACTGTTATTTGGTAAAGTAAACATATACCTTCTCAG 3839  
QY 3721 TCTATAGGATGACACCGTTGCTACGAGTGTCTGTAAGAAACAGAGAGAGATTTTA 3780  
Db 3840 TCTATAGGATGACACCGTTGCTACGAGTGTCTGTAAGAAACAGAGAGAGATTTTA 3899  
QY 3781 TTATCATTTGAAGATAGTTAAATGAATGACTGACAGTAAACAGAGTAAATTTGGCAAGCATCT 3840  
Db 3900 TTATCATTTGAAGATAGTTAAATGAATGACTGACAGTAAACAGAGTAAATTTGGCAAGCATCT 3959  
QY 3841 CAGGAACATCACCCTTAGTGAAGAAACAAATGTTGCTGACTGTTTCTTACAGTGC 3900  
Db 3960 CAGGAACATCACCCTTAGTGAAGAAACAAATGTTGCTGACTGTTTCTTACAGTGC 4019  
QY 3901 AGTGAATTTGAAGATCTGACTGCAATTAACAAACACAGAGATCTTCTGATTTGGTCT 3960  
Db 4020 AGTGAATTTGAAGATCTGACTGCAATTAACAAACACAGAGATCTTCTGATTTGGTCT 4079  
QY 3961 TCCAAACAAATGAGCATCAGTCTGAAGCCAGGAGTTGCTGATGAGCAAGAAATG 4020  
Db 4080 TCCAAACAAATGAGCATCAGTCTGAAGCCAGGAGTTGCTGATGAGCAAGAAATG 4139  
QY 4021 GTTTCAGATGATGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Db 4140 GTTTCAGATGATGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4199

QY 4081 ATGATTCAAACCTTAGTGAGACGACATCTGGGTGTGAGATGTAACAAAGCCTCTCTGAA 4140  
 |||||||  
 Db 4200 ATGGATTCAAACCTTAGTGAGACGACATCTGGGTGTGAGAGTGAACAAAGCCTCTCTGAA 4259  
 |||||||  
 QY 4141 GAGTCTCAGAGGCTTCTCTCAGAGTGATTTTAAACCCTCAGAGAGGATTCACAG 4200  
 |||||||  
 Db 4260 GACTGCTCAGAGGCTTCTCTCAGAGTGATTTTAAACCCTCAGAGAGGATTCACAG 4319  
 |||||||  
 QY 4201 CAACATAACCTGATTAACCTCCAGCAGGAAGATGGCTGAATAGAAAGCTGTAGAAACAG 4260  
 |||||||  
 Db 4320 CAACATAACCTGATTAACCTCCAGCAGGAAGATGGCTGAATAGAAAGCTGTAGAAACAG 4379  
 |||||||  
 QY 4261 CATGGAGACCAAGCCTTCTACAGCTACCTTCCATCATATAGTACTCTTCTGCTGAG 4320  
 |||||||  
 Db 4380 CATGGAGACCAAGCCTTCTACAGCTACCTTCCATCATATAGTACTCTTCTGCTGAG 4439  
 |||||||  
 QY 4321 GACCTGCCAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACTTACAGAAAACT 4380  
 |||||||  
 Db 4440 GACCTGCCAATCCAGAACAAAGCAGATCAGAAAAAGCAGTATTAACTTACAGAAAACT 4499  
 |||||||  
 QY 4381 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGCTGACAAAGTTGAGGTCT 4440  
 |||||||  
 Db 4500 AGTGAATACCTTATTAAGCCAGAAATCCAGAAAGCCTTCTGCTGACAAAGTTGAGGTCT 4559  
 |||||||  
 QY 4441 GCAGATATCTTACAGATTAATAAATAAGAACAGAGTGGAAGGTATCTCCCTCTTAA 4500  
 |||||||  
 Db 4560 GCAGATATCTTACAGATTAATAAATAAGAACAGAGTGGAAGGTATCTCCCTCTTAA 4619  
 |||||||  
 QY 4501 TGCCCATCATTAAGATGATAGGTGATGATGACAGTGTGCTGGAGTTCAGATATA 4560  
 |||||||  
 Db 4620 TGCCCATCATTAAGATGATAGGTGATGATGACAGTGTGCTGGAGTTCAGATATA 4679  
 |||||||  
 QY 4561 AACTACCCATCTCAAGAGAGAGCTCATTAAGTTGTGATGAGAGAGCAAGCTGAA 4620  
 |||||||  
 Db 4680 AACTACCCATCTCAAGAGAGAGCTCATTAAGTTGTGATGAGAGAGCAAGCTGAA 4739  
 |||||||  
 QY 4621 GAGTCTGGGCCACAGATTTTGACGGAACATCTTACTTCCCAAGCAGATCTAGAGGA 4680  
 |||||||  
 Db 4740 GAGTCTGGGCCACAGATTTTGACGGAACATCTTACTTCCCAAGCAGATCTAGAGGA 4799  
 |||||||  
 QY 4681 ACCCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
 |||||||  
 Db 4800 ACCCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4859  
 |||||||  
 QY 4741 GAAGACAGAGCCCAAGAGTCAAGCTGTGTGGCAACATACATCTTCAACCTCTGCATG 4800  
 |||||||  
 Db 4860 GAAGACAGAGCCCAAGAGTCAAGCTGTGTGGCAACATACATCTTCAACCTCTGCATG 4919  
 |||||||  
 QY 4801 AAAGTCCCAATGAAGTTGCAAGATCTGCCCCAGAGTCCAGCTGCTCTATCTACT 4860  
 |||||||  
 Db 4920 AAAGTCCCAATGAAGTTGCAAGATCTGCCCCAGAGTCCAGCTGCTCTATCTACT 4979  
 |||||||  
 QY 4861 GATACCTGCTGGTATATGCAATGGAAGAAAGTGAAGAGGAGAGAGCAAGATTTGACA 4920  
 |||||||  
 Db 4980 GATACCTGCTGGTATATGCAATGGAAGAAAGTGAAGAGGAGAGAGCAAGATTTGACA 5039  
 |||||||  
 QY 4921 GCTTCACAGAAAGGCTCAACAAAAGATGTCATGCTGTGCTGAGCCCTGACCCAGAA 4980  
 |||||||  
 Db 5040 GCTTCACAGAAAGGCTCAACAAAAGATGTCATGCTGTGCTGAGCCCTGACCCAGAA 5099  
 |||||||  
 QY 4981 GAATTTATGCTGCTGTACAAAGTTTCCAGAAAAACCCACATCTTAACTATCTAAT 5040  
 |||||||  
 Db 5100 GAATTTATGCTGCTGTACAAAGTTTCCAGAAAAACCCACATCTTAACTATCTAAT 5159  
 |||||||  
 QY 5041 ACTGAAGAGACTACTCATGTTGTATGAAGAAACAGATGCTGATTTGTGTAACGAGACA 5100  
 |||||||  
 Db 5160 ACTGAAGAGACTACTCATGTTGTATGAAGAAACAGATGCTGATTTGTGTAACGAGACA 5219  
 |||||||  
 QY 5101 CTGAATATTTTCTAGAAATTCGCGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGAC 5160  
 |||||||  
 Db 5220 CTGAATATTTTCTAGAAATTCGCGGAGAGAAATGGGTAGTATGCTATTTCTGGGTGAC 5279  
 |||||||  
 QY 5161 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGACATGATTTTGAAGTCAAGAGAGATGTG 5220  
 |||||||

Db 5280 CAGTCTATTTAAAGAAAGAAAAATGCTGAATGAGATGATTTTGAAGTCAAGAGAGATGTG 5339  
 |||||||  
 QY 5221 GTCAATGGAAGAAACACCAAGGTCCAAAGCAGAGCAAGAAATCCAGAGCAAGAAATC 5280  
 |||||||  
 Db 5340 GTCAATGGAAGAAACACCAAGGTCCAAAGCAGAGCAAGAAATCCAGAGCAAGAAATC 5339  
 |||||||  
 QY 5281 TTTCAGGGGCTAGAAATCTGTGCTATAGGGCCCTTACCAACATGCCCAGATCAATG 5340  
 |||||||  
 Db 5400 TTTCAGGGGCTAGAAATCTGTGCTATAGGGCCCTTACCAACATGCCCAGATCAATG 5459  
 |||||||  
 QY 5341 GAATGAGAGTACACTGTGTGCTCTGTGCTGTGGAAGAGCTTTCATCATCACTTACCTT 5400  
 |||||||  
 Db 5460 GAATGAGAGTACACTGTGTGCTCTGTGCTGTGGAAGAGCTTTCATCATCACTTACCTT 5519  
 |||||||  
 QY 5401 GGCACAGGTGTCACCAATGTGTGCTGACAGCCAGATGCTGAGAGAGCAATG 5460  
 |||||||  
 Db 5520 GGCACAGGTGTCACCAATGTGTGCTGACAGCCAGATGCTGAGAGAGCAATG 5579  
 |||||||  
 QY 5461 TTTCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCAGAGTGGTGTGAC 5520  
 |||||||  
 Db 5580 TTTCATGCAATTTGGCAGATGTGTGAGCACCCTGTGTGACCCAGAGTGGTGTGAC 5639  
 |||||||  
 QY 5521 AGTGTAGACACTTACCAAGTCCAGAGAGCTGACACACTACCTGATACCCAGATCCCTAC 5580  
 |||||||  
 Db 5640 AGTGTAGACACTTACCAAGTCCAGAGAGCTGACACACTACCTGATACCCAGATCCCTAC 5699  
 |||||||  
 QY 5581 AGCCACTAC 5589  
 |||||||  
 Db 5700 AGCCACTAC 5708  
 |||||||

RESULT 2  
 US-09-982-828-5  
 : Sequence 5, Application US/09982828  
 : Publication No. US20030022184A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Murphy, Patricia D.  
 : Allen, Antonette C.  
 : Alvarres, Christopher P.  
 : Critz, Brenda S.  
 : Olson, Sheri J.  
 : Thurber, Denise  
 : Zeng, Bin  
 :  
 : TITLE OF INVENTION: Coding Sequences of the Human  
 : BRC1 Gene  
 :  
 : NUMBER OF SEQUENCES: 72  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Morgan Lewis & Bockius LLP  
 : STREET: 1111 Pennsylvania Avenue N. W.  
 : CITY: Washington  
 : STATE: DC  
 : COUNTRY: USA  
 : ZIP: 20004  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/982,828  
 : FILING DATE: 22-Oct-2001  
 : CLASSIFICATION: <Unknown>  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 09/074,453  
 : FILING DATE: 1998-05-06  
 : APPLICATION NUMBER: US 08/798,691  
 : FILING DATE: 1997-02-12  
 : APPLICATION NUMBER: US 08/598,591  
 : FILING DATE: 1996-02-12  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Michael S. Tuscan  
 : REGISTRATION NUMBER: 43,210



REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAL (om13)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-982-828-5

Query Match 100.0%; Score 5587.4; DB 9; Length 5711;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGGATTATCTGCTCTTGGCGTTGAAGAAGTCAATGCTAATGCTATGAGAAA 60  
120 ATGGATTATCTGCTCTTGGCGTTGAAGAAGTCAATGCTAATGCTATGAGAAA 179  
61 ATCTTGAGTGTCCCTCTGCTGTGAGTGTATCAAGAACTGTCTCCCAAGTGTGAC 120  
180 ATCTTGAGTGTCCCTCTGCTGTGAGTGTATCAAGAACTGTCTCCCAAGTGTGAC 239  
121 CACATATTTTGCAGATTTTGCATGCTGAAGCTCTCAACAGAGAAAGAGGCTTCACAG 180  
240 CACATATTTTGCAGATTTTGCATGCTGAAGCTCTCAACAGAGAAAGAGGCTTCACAG 239  
181 TGTCTTTATGTAAGATGATATACCAAAAGAGCTCAACAGAGAAAGTACGATTTAGT 240  
300 TGTCTTTATGTAAGATGATATACCAAAAGAGCTCAACAGAGAAAGTACGATTTAGT 359  
241 CAACTGTGTGAGAGCTATTTGAAAATCATTTTGTGCTTTTACGTTTACACAGCTTTTGAG 300  
360 CAACTGTGTGAGAGCTATTTGAAAATCATTTTGTGCTTTTACGTTTACACAGCTTTTGAG 419  
301 TATGCAACAGCATATTTTGCAGAAAAGGAAATTAAGCTCTGCAACATCTAAAGAT 360  
420 TATGCAACAGCATATTTTGCAGAAAAGGAAATTAAGCTCTGCAACATCTAAAGAT 479  
361 GAAGTTTATCATCAAAAGTATGGCTACAGAAACCGTGCAGAAAGACTTTTACAGAGT 420  
480 GAAGTTTATCATCAAAAGTATGGCTACAGAAACCGTGCAGAAAGACTTTTACAGAGT 539  
421 GAACCGGAAATCCTCTCTGCGAGAAACAGCTCTAGTGTCCAACTCTTAACTTTTGA 480  
540 GAACCGGAAATCCTCTCTGCGAGAAACAGCTCTAGTGTCCAACTCTTAACTTTTGA 599  
481 ACTGTGAGAACTCTGAGACAAAGAGCGGATACAACTCTCAAAAGAGCTGTGCTCATTT 540  
600 ACTGTGAGAACTCTGAGACAAAGAGCGGATACAACTCTCAAAAGAGCTGTGCTCATTT 659  
541 GAATTTGGATCTGATTTCTCTGAAGATACGTTAATTAAGGCAACTTAATGCACTGTGGA 600  
660 GAATTTGGATCTGATTTCTCTGAAGATACGTTAATTAAGGCAACTTAATGCACTGTGGA 719  
601 GATCAAGATTTGTCAAAATCAACCTCTCAAGAAACAGGAGTGAATCACTGTGGAATTT 660  
720 GATCAAGATTTGTCAAAATCAACCTCTCAAGAAACAGGAGTGAATCACTGTGGAATTT 779  
661 GCAAAAAGGCTGCTGTGAATTTTCTGAGAGGATGTAACAAATCTGAAATCATCTCAA 720  
780 GCAAAAAGGCTGCTGTGAATTTTCTGAGAGGATGTAACAAATCTGAAATCATCTCAA 839  
721 CCCAGTAATATGATTTGAACACCACTGAGAAAGCGTGAAGGCTGAGAGGCTCAGAAAAAG 780

|||||  
840 CCCAGTAATATGATTTGAACACCACTGAGAAAGCGTGCAGCTGAGAGGCTCCAGAAAAAG 899  
781 TATCAGAGCTAGTCTGTTTCAAACTGCATGTGAGAGCATGTGGCACAATATCTCATGCC 840  
900 TATCAGAGCTAGTCTGTTTCAAACTGCATGTGAGAGCATGTGGCACAATATCTCATGCC 959  
841 AGCTCATTTACAGATGAGAACACAGCTTTTATTAAGTCACTAAAGAGAAATGATGTGAA 900  
960 AGCTCATTTACAGATGAGAACACAGCTTTTATTAAGTCACTAAAGAGAAATGATGTGAA 1019  
901 AAGCTGAATTTCTGTAATTAAGCAAAACAGCTGTGCTTACAGAGAGCCAACTATACAGA 960  
1020 AAGCTGAATTTCTGTAATTAAGCAAAACAGCTGTGCTTACAGAGAGCCAACTATACAGA 1079  
961 TGGGCTGGAAGTGAAGAAACATGTAATGATAGGCGAGCTCCCGACAGAAAAAGGTA 1020  
1080 TGGGCTGGAAGTGAAGAAACATGTAATGATAGGCGAGCTCCCGACAGAAAAAGGTA 1139  
1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATAGAGCAAACTGCTATGC 1080  
1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAGAAATGAATAGAGCAAACTGCTATGC 1199  
1081 TCAGAGAAATCTAGAGATGATGATGTTCTTGATTAACACTAAATAGCAGCAATTCAG 1140  
1200 TCAGAGAAATCTAGAGATGATGATGTTCTTGATTAACACTAAATAGCAGCAATTCAG 1259  
1141 AAGTTAATGAGTGTGTTTCCAGAACTGATGAGTGTGATGCTGTGATGCTCAGATGAT 1200  
1260 AAGTTAATGAGTGTGTTTCCAGAACTGATGAGTGTGATGCTGTGATGCTCAGATGAT 1319  
1201 GGGAGCTCTGAATCAAAATGCGCAAGTATGATGATGATGATGATGATGATGATGATGAT 1260  
1320 GGGAGCTCTGAATCAAAATGCGCAAGTATGATGATGATGATGATGATGATGATGATGAT 1379  
1380 GAATATTTCTGTTCTTACAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 1439  
1321 ATATGTAAGTGAAGAGCTTCACTCCAAATGCTGAGAGATATATGTAAGCAAAATTA 1380  
1440 ATATGTAAGTGAAGAGCTTCACTCCAAATGCTGAGAGATATATGTAAGCAAAATTA 1499  
1381 TTTGGGAAAACTATGGAAGAGGCAAGGCTCCCAACTTAAGCCATGATGATGATGAT 1440  
1500 TTTGGGAAAACTATGGAAGAGGCAAGGCTCCCAACTTAAGCCATGATGATGATGATGAT 1559  
1441 CTAAATTAAGGAGCAATTTGTTACTGAGCCACAGATATATCAAGAGGCTCCCTCACAAAT 1500  
1560 CTAAATTAAGGAGCAATTTGTTACTGAGCCACAGATATATCAAGAGGCTCCCTCACAAAT 1619  
1501 AAATTAAGGCTTAAGAGGAGACCTACATGAGGCTTCACTCTGAGAGTTTATCAAGAAA 1560  
1620 AAATTAAGGCTTAAGAGGAGACCTACATGAGGCTTCACTCTGAGAGTTTATCAAGAAA 1679  
1561 GCAGATTTGGAGATTAAGAAAGCTCTGAAATGATTAATCAAGGAGAACTCAACAAAGAG 1620  
1680 GCAGATTTGGAGATTAAGAAAGCTCTGAAATGATTAATCAAGGAGAACTCAACAAAGAG 1739  
1621 CAGAAATGCTCAAGTATGATATTAATAGTGTGATGAGAAATTAAGCAAAAGGCTGAT 1680  
1740 CAGAAATGCTCAAGTATGATATTAATAGTGTGATGAGAAATTAAGCAAAAGGCTGAT 1799  
1681 TCTATTCAAGATGAGAAAAATCTTAACCCAAATGATCACTGCAAAAAAGATGCTTTTC 1740  
1800 TCTATTCAAGATGAGAAAAATCTTAACCCAAATGATCACTGCAAAAAAGATGCTTTTC 1859  
1741 AAAAGCAAGCTGAACCTTAAGAGCAGATGATGAGAAATGAGAACTGCAATTAATATTC 1800  
1860 AAAAGCAAGCTGAACCTTAAGAGCAGATGATGAGAAATGAGAACTGCAATTAATATTC 1919  
1801 CACAAATTCAAAAGCACCTTAAGAAAGATGAGTGTGAGAGAGGCTTCTACAGAGCATATT 1860  
|||||

Db 1920 CACAATTCAAAAACCTAAAAAATAGCTGAGGAGGAGTCTTCTACAGCAGCATATT 1979  
Qy 1861 CATGGGCTTGAACTGTAAGTAATTAAGCCACCTATGTGTAAGTGA 1920  
Db 1980 CATGGGCTTGAACTGTAAGTAATTAAGCCACCTATGTGTAAGTGA 2039  
Qy 1921 ATTGATGTTCTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACATGCCAGTC 1980  
Db 2040 ATTGATGTTCTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACATGCCAGTC 2099  
Qy 1981 AGGCACAGCAGAAACCTACACTCATGGAAGTAAAGAACCTGCAACTGGACCAAGAA 2040  
Db 2100 AGGCACAGCAGAAACCTACACTCATGGAAGTAAAGAACCTGCAACTGGACCAAGAA 2159  
Qy 2041 AGTAAACAGCCAAATGAACAGACAAAGTAAAGACATGACATGATCTTCCAGAGCTG 2100  
Db 2160 AGTAAACAGCCAAATGAACAGACAAAGTAAAGACATGACATGATCTTCCAGAGCTG 2219  
Qy 2101 AAGTTAACAAATGACACCTGGTCTTTTACTAAGTGTTCAAATPCCAGTGAATTAAGAA 2160  
Db 2220 AAGTTAACAAATGACACCTGGTCTTTTACTAAGTGTTCAAATPCCAGTGAATTAAGAA 2279  
Qy 2161 TTTTGCAATCCTAGCCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGTG 2220  
Db 2280 TTTTGCAATCCTAGCCTTCCAGAGAAAGAAAGAAAGAACTAGAAACAGTTAAAGTG 2339  
Qy 2221 TCTAATTAATGCTGAAGAGCCCAAGATCTCATGTTAAGTGGAGAAAGGGTTTGCAAGCT 2280  
Db 2340 TCTAATTAATGCTGAAGAGCCCAAGATCTCATGTTAAGTGGAGAAAGGGTTTGCAAGCT 2399  
Qy 2281 GAAAGATCTGTAGAGATGACAGTATTTTCATTGTAACCTGTAAGTATGACACTCAG 2340  
Db 2400 GAAAGATCTGTAGAGATGACAGTATTTTCATTGTAACCTGTAAGTATGACACTCAG 2459  
Qy 2341 GAAATATCTCTGTTACTGGAATTTAGCACTCTAGGGAAGCAAAACAGAACCAATTA 2400  
Db 2460 GAAATATCTCTGTTACTGGAATTTAGCACTCTAGGGAAGCAAAACAGAACCAATTA 2519  
Qy 2401 TGTGTGAGTCACTGTCAGACATTTGAAAAACCCCAAGGAGCTAATTCATGTTGTTCCAA 2460  
Db 2520 TGTGTGAGTCACTGTCAGACATTTGAAAAACCCCAAGGAGCTAATTCATGTTGTTCCAA 2579  
Qy 2461 GATTAATGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAGTTAACACAGT 2520  
Db 2580 GATTAATGAAATGACACAGAGGCTTTAAGTATCCATTGGGACATGAGTTAACACAGT 2639  
Qy 2521 CGGGAACAGCATGGAATGGAAGAAATGAACTGATGATGATGAGATTA 2580  
Db 2640 CGGGAACAGCATGGAATGGAAGAAATGAACTGATGATGATGAGATTA 2699  
Qy 2581 TTCAAGGTTTCAAGCGCCAGTCATTGCTGTTTTCAAAATCCAGGAATGCGAAGAG 2640  
Db 2700 TTCAAGGTTTCAAGCGCCAGTCATTGCTGTTTTCAAAATCCAGGAATGCGAAGAG 2759  
Qy 2641 GAATGTCACACTTCTGCCCCAGTCCTGGGCTTTAAGAAACAAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTCACACTTCTGCCCCAGTCCTGGGCTTTAAGAAACAAAGTCCAAAGTCACT 2819  
Qy 2701 TTTGAATGGAACAAAGGAAGAAATCAAGGAAGAAATGAGTCTAATTCAAAGCCTGTA 2760  
Db 2820 TTTGAATGGAACAAAGGAAGAAATCAAGGAAGAAATGAGTCTAATTCAAAGCCTGTA 2879  
Qy 2761 CAGACAGTTAATATCACTGACAGCTTTCCTGTGTTGTCAGAAAGATTAAGCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCACTGACAGCTTTCCTGTGTTGTCAGAAAGATTAAGCAGTTGAT 2939  
Qy 2821 AATGCAAAATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGGC 2880  
Db 2940 AATGCAAAATGATATCAAAAGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAAGGC 2999  
Qy 2881 AAGCAAACTGACATCTACTCCAAATAACATGAGCTTTTCAAAACCCATATGCTATA 2940  
Db 3000 AAGCAAACTGACATCTACTCCAAATAACATGAGCTTTTCAAAACCCATATGCTATA 3059

Qy 2941 CCACACCTTTTCCCATCAAGTCAATTTGTTAAACCTAATGTAAGAAAAATCTGCTAGAG 3000  
Db 3060 CCACACCTTTTCCCATCAAGTCAATTTGTTAAACCTAATGTAAGAAAAATCTGCTAGAG 3119  
Qy 3001 GAAACCTTTGAGGAACATTCATGTCACCTGAAGAAAGAAATGGAACATTTCCA 3060  
Db 3120 GAAACCTTTGAGGAACATTCATGTCACCTGAAGAAAGAAATGGAACATTTCCA 3179  
Qy 3061 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAAAAATGTTTAAAGAACCCAGC 3120  
Db 3180 AGTACAGTGAACCAATTTAGCCGTAATTAACATTAAGAAAAATGTTTAAAGAACCCAGC 3239  
Qy 3121 TCAGCAATATTAATGAAGTACGTTCCAGTACTAATGAAGTGGGCTCAGTATTAAAGAA 3180  
Db 3240 TCAGCAATATTAATGAAGTACGTTCCAGTACTAATGAAGTGGGCTCAGTATTAAAGAA 3299  
Qy 3181 ATAGGTTCCAGTGAAGAAAAACATTCAGCAGAACTAGGTAGAAACAGAGGCCAAATTTG 3240  
Db 3300 ATAGGTTCCAGTGAAGAAAAACATTCAGCAGAACTAGGTAGAAACAGAGGCCAAATTTG 3359  
Qy 3241 AATGCTATGCTTGAATTAAGGCGTTTTCACACTGAGTCTTAACAAAGTCTTCGGA 3300  
Db 3360 AATGCTATGCTTGAATTAAGGCGTTTTCACACTGAGTCTTAACAAAGTCTTCGGA 3419  
Qy 3301 AGTAAATGTAACATCTGTAATTAAGAAAAAGCAAGAAATGGAAGAACTGTCAGACTGTT 3360  
Db 3420 AGTAAATGTAACATCTGTAATTAAGAAAAAGCAAGAAATGGAAGAACTGTCAGACTGTT 3479  
Qy 3361 AATACAGATTTCTCTCATATCTGATTTTCAGATACTTGAACACAGCTTATGGAAGT 3420  
Db 3480 AATACAGATTTCTCTCATATCTGATTTTCAGATACTTGAACACAGCTTATGGAAGT 3539  
Qy 3421 CATGATCTCAGGTTGTTCTGAGACCTGATGACCTGTAAGATGATGATGGAATTAAG 3480  
Db 3540 CATGATCTCAGGTTGTTCTGAGACCTGATGACCTGTAAGATGATGATGGAATTAAG 3599  
Qy 3481 GAGATACTAGTGTTCGTAAGAAATGACATTAAGAAATGCTGCTTTTGAAGCAAGC 3540  
Db 3600 GAGATACTAGTGTTCGTAAGAAATGACATTAAGAAATGCTGCTTTTGAAGCAAGC 3659  
Qy 3541 GTCCAGAAAGAGAGCTTACAGAGAGTCTAGCCCTTTACCCATACATTTGGCTCAG 3600  
Db 3660 GTCCAGAAAGAGAGCTTACAGAGAGTCTAGCCCTTTACCCATACATTTGGCTCAG 3719  
Qy 3601 GGTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAACTTCTAGTGAAGAT 3660  
Db 3720 GGTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAAAGAACTTCTAGTGAAGAT 3779  
Qy 3661 GAAGAGCTTCCCTGCTCCACACCTTGTATTTGTAAGTAACAAATATACCTTCAG 3720  
Db 3780 GAAGAGCTTCCCTGCTCCACACCTTGTATTTGTAAGTAACAAATATACCTTCAG 3839  
Qy 3721 TCTACTAGCATACAGACGTTGCTACGAGTGTCTGTAAAGAACACAGAGAAATTTA 3780  
Db 3840 TCTACTAGCATACAGACGTTGCTACGAGTGTCTGTAAAGAACACAGAGAAATTTA 3899  
Qy 3781 TTATCATTAAGAAATAGCTTAATGATGACGTAACCGAGTAATATGGAAGGCACT 3840  
Db 3900 TTATCATTAAGAAATAGCTTAATGATGACGTAACCGAGTAATATGGAAGGCACT 3959  
Qy 3841 CAGGAACATCACCTTATGAGAGAAACAAATGTTCTGATCTGTTTTCACAGTTC 3900  
Db 3960 CAGGAACATCACCTTATGAGAGAAACAAATGTTCTGATCTGTTTTCACAGTTC 4019  
Qy 3901 ACTGAATGGAAGACTTGAATCAATCAAAACCCAGAGTCTCTTCTGATGATGTTCT 3960  
Db 4020 ACTGAATGGAAGACTTGAATCAATCAAAACCCAGAGTCTCTTCTGATGATGTTCT 4079  
Qy 3961 TCCAAACAAATGAGCATACGTCGAAGCCAGGAGTGTGTCGAGGCAAGAAATG 4020  
Db 4080 TCCAAACAAATGAGCATACGTCGAAGCCAGGAGTGTGTCGAGGCAAGAAATG 4139

QY	4021	GTTCACAGTATGTAAGAAAGGAAAGCGGCTTGGAAAGAAATATATCAAGAAAGACCAAGC	4080
Db	4140	GTTCACAGTATGTAAGAAAGGAAAGCGGCTTGGAAAGAAATATATCAAGAAAGACCAAGC	4199
QY	4081	ATGATTCCAACTTAGGTGAAGCAGCATCTGGGTGTGAGAGTGAAGAACGGCTCTCGAA	4140
Db	4200	ATGATTCCAACTTAGGTGAAGCAGCATCTGGGTGTGAGAGTGAAGAACGGCTCTCGAA	4259
QY	4141	GACTGCTAAGGGCTATCTCTCAGAGTACATTTTAACTACAGAGAGGATACCATG	4200
Db	4260	GACTGCTAAGGGCTATCTCTCAGAGTACATTTTAACTACAGAGAGGATACCATG	4319
QY	4201	CAACATACTCGATTAAGCTCCAGAGGAATGCGTGAACATAAGAGCGTGTGAACAG	4260
Db	4320	CAACATACTCGATTAAGCTCCAGAGGAATGCGTGAACATAAGAGCGTGTGTGAACAG	4379
QY	4261	CATGGAGACCCAGCCTTCTTACAGAGTACCTTCCATCATAACTGACTCTTGCCCTTGAG	4320
Db	4380	CATGGAGACCCAGCCTTCTTACAGAGTACCTTCCATCATAACTGACTCTTGCCCTTGAG	4439
QY	4331	GACCTGGCAATCCAGAACCAAGCACAATCAGAAAAAGCAGTATTACTTCAGAAAAAGT	4380
Db	4440	GACCTGGCAATCCAGAACCAAGCACAATCAGAAAAAGCAGTATTACTTCAGAAAAAGT	4499
QY	4381	AGTGAATACCTCTATTAAGCCCAAAATCCAGAAAGCCCTTCTGCTGCAAGTTGAGAGTCT	4440
Db	4500	AGTGAATACCTCTATTAAGCCCAAAATCCAGAAAGCCCTTCTGCTGCAAGTTGAGAGTCT	4559
QY	4441	GCAGATAGTTTACACAGTAAATAAAGAAACAGAGAGGAGAAAGTCATCCCTCTTCAA	4500
Db	4560	GCAGATAGTTTACACAGTAAATAAAGAAACAGAGAGGAGAAAGTCATCCCTCTTCAA	4619
QY	4501	TGCCCATCATTAGATGATAGGTGGTACATGACAGATTGCTCTGGAGTCTTCAGAAATAGA	4560
Db	4620	TGCCCATCATTAGATGATAGGTGGTACATGACAGATTGCTCTGGAGTCTTCAGAAATAGA	4679
QY	4561	AACATACCCATCTCAAGAGAGAGCTCTATTAAGTTTGTATGTGGAGAGCAACAGCTGGAA	4620
Db	4680	AACATACCCATCTCAAGAGAGAGCTCTATTAAGTTTGTATGTGGAGAGCAACAGCTGGAA	4739
QY	4621	GAGTCTGGGGCACACAGATTTGACGAAACAATCTTACTTGCCCAAGCAAGATCTGAGAGGA	4680
Db	4740	GAGTCTGGGGCACACAGATTTGACGAAACAATCTTACTTGCCCAAGCAAGATCTGAGAGGA	4799
QY	4681	ACCCTTAACCTGGAATCTGGAATCAGGCTCTTCTGATGACCCCTGGAATCTGATCCTCT	4740
Db	4800	ACCCTTAACCTGGAATCTGGAATCAGGCTCTTCTGATGACCCCTGGAATCTGATCCTCT	4859
QY	4741	GAACACAGAGCCCCAGAGTCAAGCTCGTGTGGCAACATACATCTTCAACTCTGCATTG	4800
Db	4860	GAACACAGAGCCCCAGAGTCAAGCTCGTGTGGCAACATACATCTTCAACTCTGCATTG	4919
QY	4801	AAAGTCCCCCAATTTGAAGATGACGAATCTGCCAGAGTCCAGCTGCGTGCATACACT	4860
Db	4920	AAAGTCCCCCAATTTGAAGATGACGAATCTGCCAGAGTCCAGCTGCGTGCATACACT	4979
QY	4861	GATACGCTGGGTATAATGCAATGGAAGAAAGTGTGAGCAGGAGGAAGCCAGAAATTGACA	4920
Db	4980	GATACGCTGGGTATAATGCAATGGAAGAAAGTGTGAGCAGGAGGAAGCCAGAAATTGACA	5039
QY	4921	GCTTCACAGAAAGGGTCAACAAAAGATGTCATGCTGTCTGCGCTGACCCAGAA	4980
Db	5040	GCTTCACAGAAAGGGTCAACAAAAGATGTCATGCTGTCTGCGCTGACCCAGAA	5099
QY	4981	GAATTTATGCTCGTGTCACAGTTTGCCAGAAACACACATCACTTTAACTAACTAATT	5040
Db	5100	GAATTTATGCTCGTGTCACAGTTTGCCAGAAACACACATCACTTTAACTAACTAATT	5159
QY	5041	ACTGAGAGACTACTCATGTTGTTATGAAAAACAGATCCTGAGTTGTGTGGAAGGACA	5100
Db	5160	ACTGAGAGACTACTCATGTTGTTATGAAAAACAGATCCTGAGTTGTGTGGAAGGACA	5219
QY	5101	CTGAAATATTTTCTAGGAATTCGGGAGAGAAATGGGTAGTTAGCTATTTCTGGGTGACC	5160

Db	5220	CTGAAATATTTTCTAGGAAATGCGGAGAGAAATGGTAGTTACTATTTCTGGGTGACC	5279
Qy	5161	CAGTCATATTTAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG	5220
Db	5280	CAGCTCATTTAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGAGATGTG	5339
Qy	5221	GTCATATGGAAGAAACCACCAAGTGTCCAAAGGAGACAGAGAAATCCACAGAAACATC	5280
Db	5340	GTCATATGGAAGAAACCACCAAGTGTCCAAAGGAGACAGAGAAATCCACAGAAACATC	5399
Qy	5281	TTCAAGGGGGCTAGAAATCTGTGTATGAGGGCCCTTCAACAATGCCACAGATCAACTG	5340
Db	5400	TTCAAGGGGGCTAGAAATCTGTGTATGAGGGCCCTTCAACAATGCCACAGATCAACTG	5459
Qy	5341	GAATGGATGGTACAGCTGTGTGTGTCCTGTGTGTGAAGGAGCTTTTCATCATTCACCTT	5400
Db	5460	GAATGGATGGTACAGCTGTGTGTGTCCTGTGTGTGAAGGAGCTTTTCATCATTCACCTT	5519
Qy	5401	GGCAGAGGTGCCACCCCAATGTTCGGTGTGTGAGCAATGGCTGTGCAGAGAGCAATGGC	5460
Db	5520	GGCAGAGGTGCCACCCCAATGTTCGGTGTGTGAGCAATGGCTGTGCAGAGAGCAATGGC	5579
Qy	5461	TTTCATGCAATTTGGGAGAGATGTGTGAGGCAACCTGTGTGACCCGAGAGTGGGTGTGGAC	5520
Db	5580	TTTCATGCAATTTGGGAGAGATGTGTGAGGCAACCTGTGTGACCCGAGAGTGGGTGTGGAC	5639
Qy	5521	AGTGTAGACCTCTACCAAGTGGCAGAGACTGGACACCTACCTGATATCCCAAGATCCCCAC	5580
Db	5640	AGTGTAGACCTCTACCAAGTGGCAGAGACTGGACACCTACCTGATATCCCAAGATCCCCAC	5699
Qy	5581	AGGCACCTAC 5589	
Db	5700	AGCCACTAC 5708	

RESULT 3  
 US-09-734-672-5  
 Sequence 5, Application US/09734672  
 Publication No. US20020183268A1  
 GENERAL INFORMATION:  
 APPLICANT: Murphy, Patricia D.  
 Allen, Antonette C.  
 Alvarez, Christopher P.  
 Critz, Brenda S.  
 Olson, Sheri J.  
 Schelter, Denise B.  
 Zeng, Bin  
 TITLE OF INVENTION: Coding Sequences of the Human  
 BRCA1 Gene  
 NUMBER OF SEQUENCES: 72  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morgan Lewis & Bockius LLP  
 STREET: 1111 Pennsylvania Ave., N.W.  
 CITY: Washington  
 STATE: District of Columbia  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/734,672  
 FILING DATE: 03-Dec-2000  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/966,436  
 FILING DATE: 07-No. US20020183268A1-97  
 APPLICATION NUMBER: US 08/598,591  
 FILING DATE: 12-Feb-96  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michael S. Tuscan

REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCA1  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-734-672-5

Query Match 99.9% Score 5582.6 DB 9 Length 5711:

Best Local Similarity 99.9% Pred. No. 0:  
Matches 5585: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGATTTATCTGCTTCGCGTTGAGAGATACAAATGTCATTATGCTATGCAGAA 60  
DB 120 ATGATTTATCTGCTTCGCGTTGAGAGATACAAATGTCATTATGCTATGCAGAA 179  
QY 61 ATCTTAGAGTGTCCCATCTGCTGAGGTGATCAGAGAACTGTCTCCCAAGTGTGAC 120  
DB 180 ATCTTAGAGTGTCCCATCTGCTGAGGTGATCAGAGAACTGTCTCCCAAGTGTGAC 239  
QY 121 CACATATTTTGCAGATTTTGCATGCGAATCTCTCAACAGAGAGAGAGGCTTCACAG 180  
DB 240 CACATATTTTGCAGATTTTGCATGCGAATCTCTCAACAGAGAGAGAGGCTTCACAG 299  
QY 181 TGTCTTTATGTAAGATGATTAACCAAGAGAGGCTTACAGAGAGAGATTTAGT 240  
DB 300 TGTCTTTATGTAAGATGATTAACCAAGAGAGGCTTACAGAGAGAGATTTAGT 359  
QY 241 CAAGTGTGTAAGAGTATGTAAGATGATTTGCTTTTACAGTTCACAGAGTGGAG 300  
DB 360 CAAGTGTGTAAGAGTATGTAAGATGATTTGCTTTTACAGTTCACAGAGTGGAG 419  
QY 301 TATGCAACAGCTATATATTTTGCAGAGAGAGAGAGAGAGTCTCTGTAACATCAAGAT 360  
DB 420 TATGCAACAGCTATATATTTTGCAGAGAGAGAGAGAGTCTCTGTAACATCAAGAT 479  
QY 361 GAAGTGTCTATCAATCAAGATATGAGGCTACAGAGAGAGGCTTACAGAGT 420  
DB 480 GAAGTGTCTATCAATCAAGATATGAGGCTACAGAGAGAGGCTTACAGAGT 539  
QY 421 GAAGCCGAGAGAGTCTCTGAGAGAGAGAGAGAGAGTCTCTTCAACCTTGA 480  
DB 540 GAAGCCGAGAGAGTCTCTGAGAGAGAGAGAGAGTCTCTTCAACCTTGA 599  
QY 481 ACTGTGAGAGTCTGAGAGAGAGAGAGAGAGTCTCTTCAACCTTGA 540  
DB 600 ACTGTGAGAGTCTGAGAGAGAGAGAGAGTCTCTTCAACCTTGA 659  
QY 541 GAATTTGGATCTGATTTCTTGAAGATACCGTTAATAGGCAACTTATGCGAGTGGGA 600  
DB 660 GAATTTGGATCTGATTTCTTGAAGATACCGTTAATAGGCAACTTATGCGAGTGGGA 719  
QY 601 GATCAGAGATTTGTAACAATCACCCCTCAAGAGAGAGAGAGATGATTTGATTTCT 660  
DB 720 GATCAGAGATTTGTAACAATCACCCCTCAAGAGAGAGAGATGATTTGATTTCT 779  
QY 661 GCAAAAAAGGCTGTTGTAATTTCTGAGAGAGATGTAACAATTAAGATCATATCAA 720  
DB 780 GCAAAAAAGGCTGTTGTAATTTCTGAGAGAGATGTAACAATTAAGATCATATCAA 839

QY 721 CCCAGTAATATGATTTTGAACACACTGAGAGAGGCTGAGAGGATCCAGAAAG 780  
DB 840 CCCAGTAATATGATTTTGAACACACTGAGAGAGGCTGAGAGGATCCAGAAAG 899  
QY 781 TATCAGAGATTTCTTTTCAAACTGTGAGAGAGAGAGTGGCAGAAATATCATGCC 840  
DB 900 TATCAGAGATTTCTTTTCAAACTGTGAGAGAGAGAGTGGCAGAAATATCATGCC 959  
QY 841 AGCTCATTCAGAGATGAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 900  
DB 960 AGCTCATTCAGAGATGAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1019  
QY 901 AAGGCTGAATTTCTGTAATATTAAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 960  
DB 1020 AAGGCTGAATTTCTGTAATATTAAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1079  
QY 961 TGGGCTGAGAGATGAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1020  
DB 1080 TGGGCTGAGAGATGAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1139  
QY 1021 GATCTGAATGCTGATCCCTGCTGAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1080  
DB 1140 GATCTGAATGCTGATCCCTGCTGAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1199  
QY 1081 TCAGAGATTCCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
DB 1200 TCAGAGATTCCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1259  
QY 1141 AAGGTAATGAGTGTGTTTCCAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1200  
DB 1260 AAGGTAATGAGTGTGTTTCCAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1319  
QY 1201 GGGAGTCTGATTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1260  
DB 1320 GGGAGTCTGATTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1379  
QY 1261 GAATATTCGTTCTTCAAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1320  
DB 1380 GAATATTCGTTCTTCAAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1439  
QY 1321 ATATGTAAGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1380  
DB 1440 ATATGTAAGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1499  
QY 1381 TTTGGAGAGAGTATTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1440  
DB 1500 TTTGGAGAGAGTATTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGATGATTTAG 1559  
QY 1441 CTAATTTATGAGAGATTTCTTCAAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1500  
DB 1560 CTAATTTATGAGAGATTTCTTCAAGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1619  
QY 1501 AAATTTAAGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1560  
DB 1620 AAATTTAAGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1679  
QY 1561 GCAGATTTGGAGTTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGAGATGATTTAG 1620  
DB 1680 GCAGATTTGGAGTTCAGAGAGAGAGAGAGTATTAATGAGAGAGAGAGATGATTTAG 1739  
QY 1621 CAGATTTGATGAGATTTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1680  
DB 1740 CAGATTTGATGAGATTTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1799  
QY 1681 TCTATTTGAGATGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAT 1740  
DB 1800 TCTATTTGAGATGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAT 1859  
QY 1741 AAAAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAGATGATTTAG 1800  
DB 1860 AAAAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAGATGATTTAG 1919  
QY 1801 CACAATTCAGAGAGAGAGAGTATTAATGAGAGAGAGAGAGTATTAATGAGAGAGAGAT 1860

Db	1920	CACAAATTCMAAAGACACCTTAATAAATAATAGCTGAGGAGAAAGTCTTCTACAGGCATATT	1979
QY	1861	CATCGCTTGAAGTACTAGTACGTAGAAATCTAAGCCACCTAAATTGTACTGAATTGCAA	1920
Db	1980	CATCGCTGTGAAGTACTAGTACGTAGAAATCTAAGCCACCTAAATTGTACTGAATTGCAA	2039
QY	1921	ATTGTATAGTGTCTTACGACGTGAAGAGATTAAGAAGAAAAAGTCAACCAATGCCAGTC	1980
Db	2040	ATTGTATAGTGTCTTACGACGTGAAGAGATTAAGAAGAAAAAGTCAACCAATGCCAGTC	2099
QY	1981	AGGCACAGCAAAAACCTCAACCTCATGGAAGGTAAAGAACTGGCAACTGGAGCCAAAGAG	2040
Db	2100	AGGCACAGCAAAAACCTCAACCTCATGGAAGGTAAAGAACTGGCAACTGGAGCCAAAGAG	2159
QY	2041	AGTAACAAGCCMAATGAAACAGACAAAGTAAAGACATGACAGTGTACTTTCCAGAGCTG	2100
Db	2160	AGTAACAAGCCMAATGAAACAGACAAAGTAAAGACATGACAGTGTACTTTCCAGAGCTG	2219
QY	2101	AAGTTAAACAATGCACCTGGTTCCTTTACTAAGTGTTCAAATACCAAGTGAACCTTAAAGAA	2160
Db	2220	AAGTTAAACAATGCACCTGGTTCCTTTACTAAGTGTTCAAATACCAAGTGAACCTTAAAGAA	2279
QY	2161	TTTGTCAATCCTACCTCTTCCAAAGAGAAAGAAAGAAACCTAAGAAACAGTAAAGTG	2220
Db	2280	TTTGTCAATCCTACCTCTTCCAAAGAGAAAGAAAGAAACCTAAGAAACAGTAAAGTG	2339
QY	2221	TCTATATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGCAAACT	2280
Db	2340	TCTATATAATGCTGAAGACCCCAAGATCTCATGTTAAGTGAAGAAAGGTTTGCAAACT	2399
QY	2281	GAAAGATCTGTAGAGAGTAGCAGTATTTTCATTTGTACTCGTACTGATTAATGGCAGTCAG	2340
Db	2400	GAAAGATCTGTAGAGAGTAGCAGTATTTTCATTTGTACTCGTACTGATTAATGGCAGTCAG	2459
QY	2341	GAAAGTATCTCGTACTGAGAGTGAAGTGAACACTTAAAGGAAAGCAAGAAACCAATATAA	2400
Db	2460	GAAAGTATCTCGTACTGAGAGTGAAGTGAACACTTAAAGGAAAGCAAGAAACCAATATAA	2519
QY	2401	TGTGTGATCGTCAAGTGTGACGATTTTGAAACCCCAAGGACTAATTCATGTGTGTTCCAA	2460
Db	2520	TGTGTGATCGTCAAGTGTGACGATTTTGAAACCCCAAGGACTAATTCATGTGTGTTCCAA	2579
QY	2461	GATTAATAGAAATGACACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT	2520
Db	2580	GATTAATAGAAATGACACAGAGAGGCTTTAAGTATCCATTGGGACATGAAGTTAACACAGT	2639
QY	2521	CGGGAACACAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTCACAAATACA	2580
Db	2640	CGGGAACACACAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTCACAAATACA	2699
QY	2581	TTTCAAGGTTTCAAAAGCGGCAGTCAATTTGGCTGTGTTTCAATTCACAGAAATGCGAAGAG	2640
Db	2700	TTTCAAGGTTTCAAAAGCGGCAGTCAATTTGGCTGTGTTTCAATTCACAGAAATGCGAAGAG	2759
QY	2641	GAATGTGCACATTTCTGTGCCACTCTGGGTCCTTTAAAGAACCAAGTCCAAAAGTCACT	2700
Db	2760	GAATGTGCACATTTCTGTGCCACTCTGGGTCCTTTAAAGAACCAAGTCCAAAAGTCACT	2819
QY	2701	TTTGAATGTGAACAAAAGGAAGAAATCAAGGAAGAAATGAGTCTAATATCAAGCCTGTA	2760
Db	2820	TTTGAATGTGAACAAAAGGAAGAAATCAAGGAAGAAATGAGTCTAATATCAAGCCTGTA	2879
QY	2761	CAGACAGTAAATATACATGCAGAGCTTTTCGTGTGTGGTGCAGAAAGATAAAGCCAGTTGAT	2820
Db	2880	CAGACAGTAAATATACATGCAGAGCTTTTCGTGTGTGGTGCAGAAAGATAAAGCCAGTTGAT	2939
QY	2821	AATGCCAAATGTAGTATCAAAAGAGGCTCTAGTTTTGTCTATCATCTCAAGTGAAGGC	2880
Db	2940	AATGCCAAATGTAGTATCAAAAGAGGCTCTAGTTTTGTCTATCATCTCAAGTGAAGGC	2999
QY	2881	AACGAACCTGAGCTCATTTACTCCAATTAACATGAGACTTTTACAAAACCCATATCGATA	2940

Dp	3000	AACGAAACTGGACATCTATTCACGCAAAATTAACATGSCACTTTTACAAACCAACCCATATCGTATA	305
Oy	2941	CCACCACCTTTTCCCATCAAGTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG	3000
Dp	3060	CCACCACCTTTTCCCATCAAGTCATTTGTAAACTAAATGTAGAAAAATCTGCTAGAG	3119
Oy	3001	GAAAACCTTTGGGAAACATTCANATGTCACCTGAAAGAAATGGGAAATGACAACTTCCA	3066
Dp	3120	GA AAAACCTTTGGGAAACATTCANATGTCACCTGAAAGAAATGGGAAATGACAACTTCCA	3179
Oy	3061	AGTACAGAGCACACAATTAGCCGTATATACATTTAGAGAAATGTGTTTAAAGAACCCAGC	3120
Dp	3180	AGTACAGAGCACACAATTAGCCGTATATACATTTAGAGAAATGTGTTTAAAGAACCCAGC	3239
Oy	3121	TC AAGCAATATTAATGAAGTAGTTCACGTACTAATGAAGTGGGCTCCAGTATTAAATGA	3180
Dp	3240	TC AAGCAATATTAATGAAGTAGTTCACGTACTAATGAAGTGGGCTCCAGTATTAAATGA	3299
Oy	3181	ATAGGTTCCAGTATGA A A A C A T T C A G C A G A C T A G T A G A A A C A G A G C C A A A T T G	3240
Dp	3300	ATAGGTTCCAGTATGA A A A C A T T C A G C A G A C T A G T A G A A A C A G A G C C A A A T T G	3359
Oy	3241	AATGCTATGCTAGCATTTAGGGGTTTTGGACCTGAGGCTCTTAAACAAGCTTCCTGGA	3300
Dp	3360	AATGCTATGCTAGCATTTAGGGGTTTTGGACCTGAGGCTCTTAAACAAGCTTCCTGGA	3419
Oy	3301	AGTATTTGTAAGCATCCCTGAATATAAAAGCAGAGATATGAAGAATAGTTACAGACTGT	3360
Dp	3420	AGTATTTGTAAGCATCCCTGAATATAAAAGCAGAGATATGAAGAATAGTTACAGACTGT	3479
Oy	3361	AATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGGAATAGT	3420
Dp	3480	AATACAGATTTCTCTCCATATCTGATTTCAATTAACATAGAACAGCCTATGGGAATAGT	3539
Oy	3421	CATGCAATCTCAGGTTTGTCTGAGACACACTGATGACCTGTATGATGATGGTGAATTAAG	3480
Dp	3540	CATGCAATCTCAGGTTTGTCTGAGACACACTGATGACCTGTATGATGATGGTGAATTAAG	3599
Oy	3481	GAAGTACTACTTTTGTCTGA A A A A T G A C A T T T A A G A A A T T C T G C T T T T A G C A A A A G C	3540
Dp	3600	GAAGTACTACTTTTGTCTGA A A A A T G A C A T T T A A G A A A T T C T G C T T T T A G C A A A A G C	3655
Oy	3541	GTCCAGAAAGAGAGCTTAGCAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAG	3600
Dp	3660	GTCCAGAGAGAGAGCTTAGCAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAG	3719
Oy	3601	GGTTACCGAAGAGGGGCCAACAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT	3660
Dp	3720	GGTTACCGAAGAGGGGCCAACAATTAGAGTCTCAGAGAGAACTTATCTAGTGAAGAT	3779
Oy	3661	GAAGAGCTTCCCTGCTTCCAAACACTGTATTGGTAAAGTAAACAATATACCTTCTCAG	3720
Dp	3780	GAAGAGCTTCCCTGCTTCCAAACACTGTATTGGTAAAGTAAACAATATACCTTCTCAG	3839
Oy	3721	TCTACTAGGCACTAGCACCGTTGCTACCGAGTGTGTTTAAGAACACAGAGAGAAATTA	3780
Dp	3840	TCTACTAGGCACTAGCACCGTTGCTACCGAGTGTGTTTAAGAACACAGAGAGAAATTA	3899
Oy	3781	TTATCATTTGAAGAAATAGCTTAATATGACTGCAGTAAACAGATATATTTGGCAAGGACATCT	3840
Dp	3900	TTATCATTTGAAGAAATAGCTTAATATGACTGCAGTAAACAGATATATTTGGCAAGGACATCT	3959
Oy	3841	CAGGAACATCACTTAGTGAAGGAAACAAATGTGTTCTGTACTGTGTTTCTTCCACAGTGC	3900
Dp	3960	CAGGAACATCACTTAGTGAAGGAAACAAATGTGTTCTGTACTGTGTTTCTTCCACAGTGC	4019
Oy	3901	AGTGAATTTGGAAGACTGATGCAAAATTAACAACACCAAGATGCTTTCTTATTTGGTTCT	3960
Dp	4020	AGTGAATTTGGAAGACTGATGCAAAATTAACAACACCAAGATGCTTTCTTATTTGGTTCT	4079
Oy	3961	TC CA A A C A A A T G A G G C A T C A G T C T G A A A C C C A G G A G T T G C T G A G T G A C A A G A A T T G	4020
Dp	4080	TC CA A A C A A A T G A G G C A T C A G T C T G A A A C C C A G G A G T T G C T G A G T G A C A A G A A T T G	4139

4021 GTTTCAGATGATGAAGAAGAGGAGCGGCTTGGAAAGAAATATCAAGAGCAAGC 4080  
|||||  
4140 GTTTCAGATGATGAAGAAGAGGAGCGGCTTGGAAAGAAATATCAAGAGCAAGC 4199  
|||||  
4081 ATGGAATTCAACTTAGGTGAAGCAGCATCGGCTGTGAGAGTGAACACAGCGTCTGAA 4140  
|||||  
4200 ATGGAATTCAACTTAGGTGAAGCAGCATCGGCTGTGAGAGTGAACACAGCGTCTGAA 4259  
|||||  
4141 GACTGCTCAGGCGTATCCTCTCAGAGTGAATTTTAAACCCTCAGCAGAGGATACCATG 4200  
|||||  
4260 GACTGCTCAGGCGTATCCTCTCAGAGTGAATTTTAAACCCTCAGCAGAGGATACCATG 4319  
|||||  
4201 CAACATTAACCTGATTAAGCTCCAGCAGAAATGCTGAATAGAGCTGTGTAAGACAG 4260  
|||||  
4320 CAACATTAACCTGATTAAGCTCCAGCAGAAATGCTGAATAGAGCTGTGTAAGACAG 4379  
|||||  
4261 CATGGAGGCGAGCGCTTCTAAGCAGCTACCTTCCATCATTAAGTACTCTTCCCTTGAG 4320  
|||||  
4380 CATGGAGGCGAGCGCTTCTAAGCAGCTACCTTCCATCATTAAGTACTCTTCCCTTGAG 4439  
|||||  
4321 GACCTGCGCAATTCACAAAGCAGATCGAAGAAAGCAATTTACTTCAAGAAAGT 4380  
|||||  
4440 GACCTGCGCAATTCACAAAGCAGATCGAAGAAAGCAATTTACTTCAAGAAAGT 4499  
|||||  
4381 AGTGAATACCTTATTAAGCAGATCGAAGGCGCTTCTGCTGACAAAGTTGAGAGTCT 4440  
|||||  
4500 AGTGAATACCTTATTAAGCAGATCGAAGGCGCTTCTGCTGACAAAGTTGAGAGTCT 4559  
|||||  
4441 GCAGATAGTCTTACCAAGTAAATTAAGAACCAAGAGTGAAGAGTCAATCCCTTTGAA 4500  
|||||  
4560 GCAGATAGTCTTACCAAGTAAATTAAGAACCAAGAGTGAAGAGTCAATCCCTTTGAA 4619  
|||||  
4501 TGCCCATCATTTAGATAGTGTGTACATGACAGATGCTCTGGAGTCTTCAGAAATAGA 4560  
|||||  
4620 TGCCCATCATTTAGATAGTGTGTACATGACAGATGCTCTGGAGTCTTCAGAAATAGA 4679  
|||||  
4561 AACTACCATCTTCAAGAGGAGCTCATTAAGTTGTGATGTGAGAGCAACAGCTGGA 4620  
|||||  
4680 AACTACCATCTTCAAGAGGAGCTCATTAAGTTGTGATGTGAGAGCAACAGCTGGA 4739  
|||||  
4621 GAGTGTGGGCGCACAGATTTGACAGGAACATCTTACTTCCCAAGCAGATCTAGAGGA 4680  
|||||  
4740 GAGTGTGGGCGCACAGATTTGACAGGAACATCTTACTTCCCAAGCAGATCTAGAGGA 4799  
|||||  
4681 ACCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGAGCCCTGATGATGATCTCT 4740  
|||||  
4800 ACCCCTTACCTGGAATCTGGAATCAGCTCTTCTGATGAGCCCTGATGATGATCTCTCT 4859  
|||||  
4741 GAAGACAGAGCCCGACAGTCAAGCTGTGTTGGCAACATACATCTTCAACCTCTGATG 4800  
|||||  
4860 GAAGACAGAGCCCGACAGTCAAGCTGTGTTGGCAACATACATCTTCAACCTCTGATG 4919  
|||||  
4801 AAAGTTCCCAATTTGAAGTTGACAGATCTGCCAGAGTCCAGCTGCTCTACTACT 4860  
|||||  
4920 AAAGTTCCCAATTTGAAGTTGACAGATCTGCCAGAGTCCAGCTGCTCTACTACT 4979  
|||||  
4861 GATACGCTGGGTATTAATGCAATGGAAGAGTGAAGAGGAGAGCAAGCAAGATGACA 4920  
|||||  
4980 GATACGCTGGGTATTAATGCAATGGAAGAGTGAAGAGGAGAGCAAGCAAGATGACA 5039  
|||||  
4921 GCTTCACAGAAAGGTTCAACAAAGAAATGCTGATGCTGCTGACCTGACCCAGAA 4980  
|||||  
5040 GCTTCACAGAAAGGTTCAACAAAGAAATGCTGATGCTGCTGACCTGACCCAGAA 5099  
|||||  
4981 GAATTTATGCTGTGTGTAAGTTGCGCAGAAACACACATCATCTTAACTAATCTAAT 5040  
|||||  
5100 GAATTTATGCTGTGTGTAAGTTGCGCAGAAACACACATCATCTTAACTAATCTAAT 5159  
|||||  
5041 ACTGAAGAGACTACTCATGTGTGTATGAAGACAGATGAGTGTGTGTGAAGGACA 5100  
|||||  
5160 ACTGAAGAGACTACTCATGTGTGTATGAAGACAGATGAGTGTGTGTGAAGGACA 5219  
|||||

QY 5101 CTGAATATTTTCTAGGAATTCGGGAGGAGAAATGGGTAGTAACTATTTCTGGGTGAC 5160  
|||||  
Db 5220 CTGAATATTTTCTAGGAATTCGGGAGGAGAAATGGGTAGTAACTATTTCTGGGTGAC 5279  
|||||  
QY 5161 CAGCTATTTAAGAAAGAAATATGCTGAATGACATGATTTTGAATCAGAGAGATGTG 5220  
|||||  
Db 5280 CAGCTATTTAAGAAAGAAATATGCTGAATGACATGATTTTGAATCAGAGAGATGTG 5339  
|||||  
QY 5221 GTCAATGGAAGAAACCAAGGCTCCAAAGCCGAGCAAGAAATCCAGACAGAAAGATC 5280  
|||||  
Db 5340 GTCAATGGAAGAAACCAAGGCTCCAAAGCCGAGCAAGAAATCCAGACAGAAAGATC 5399  
|||||  
QY 5281 TTCAGGGGCTTACAAATCTGTCTATGGGCCCTTACCAACATGCCACAGATCAATG 5340  
|||||  
Db 5400 TTCAGGGGCTTACAAATCTGTCTATGGGCCCTTACCAACATGCCACAGATCAATG 5459  
|||||  
QY 5341 GAATGATGATACAGCTGTGTGTGCTTCTGTGTGAAGAGCTTTCATCATTCACCTT 5400  
|||||  
Db 5460 GAATGATGATGATACAGCTGTGTGTGCTTCTGTGTGAAGAGCTTTCATCATTCACCTT 5519  
|||||  
QY 5401 GGCACAGGCTCCACCCTTGTGTGTGTCAGCCGAGATGCTTGACAGACAGCAATGGC 5460  
|||||  
Db 5520 GGCACAGGCTCCACCCTTGTGTGTGTCAGCCGAGATGCTTGACAGACAGCAATGGC 5579  
|||||  
QY 5461 TTCATGCAATTTGGCAGATGTGTGAGCAGCAGCTGTGTGACCCGAGAGTGGGTGGAC 5520  
|||||  
Db 5580 TTCATGCAATTTGGCAGATGTGTGAGCAGCAGCTGTGTGACCCGAGAGTGGGTGGAC 5639  
|||||  
QY 5521 AGTGTAGCACTTACCAAGTGGCAGAGCTGACACCTTACTGATACCCAGATCCCCAC 5580  
|||||  
Db 5640 AGTGTAGCACTTACCAAGTGGCAGAGCTGACACCTTACTGATACCCAGATCCCCAC 5699  
|||||  
QY 5581 AGCCACTAC 5589  
|||||  
Db 5700 AGCCACTAC 5708  
|||||

RESULT 4  
US-09-982-828-3  
; Sequence 3, Application US/09982828  
; Publication NO. US20030022184A1  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Patricia D.  
; Allen, Antonette C.  
; Alvarez, Christopher P.  
; Critz, Brenda S.  
; Olson, Sheri J.  
; Thurber, Denise  
; Zeng, Bin  
; TITLE OF INVENTION: Coding Sequences of the Human  
; BRCAL Gene  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morgan Lewis & Bockius LLP  
; STREET: 1111 Pennsylvania Avenue N. W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/982,828  
; FILING DATE: 22-Oct-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/074,453  
; FILING DATE: 1998-05-06  
; APPLICATION NUMBER: US 08/798,691  
; FILING DATE: 1997-02-12

APPLICATION NUMBER: US 08/598,591  
FILING DATE: 1996-02-12  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5053-01-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3001  
TELEFAX: 202-739-3000  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAL (cm12)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-982-828-3

Query Match 99.9%; Score 5582.6; DB 9; Length 5711;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 5585; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ATGATTATTCCTGCTCCGCTTGGAAGATCAAAATGTCATTAACTGATGACAGAA 60  
DB 120 ATGATTATTCCTGCTCCGCTTGGAAGATCAAAATGTCATTAACTGATGACAGAA 179  
OY 61 ATCTAGAGTCCCATCTGCTGAGTGTATGATCAAGAACCTGCTCCCAAAAGTGCAC 120  
DB 180 ATCTAGAGTCCCATCTGCTGAGTGTATGATCAAGAACCTGCTCCCAAAAGTGCAC 239  
OY 121 CACATATTTTGCAAAATTTGCATGCTGAACTTCTCAACAGAAAGAGGCCCTTCACAG 180  
DB 240 CACATATTTTGCAAAATTTGCATGCTGAACTTCTCAACAGAAAGAGGCCCTTCACAG 299  
OY 181 TGTCTTTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGATGAGATTAGT 240  
DB 300 TGTCTTTTATGTAAGATGATATACCAAAAGAGCCTACAGAAAGATGAGATTAGT 359  
OY 241 CAACTGTTGAAGAGCTATGAAATATGATTTGCTCTTTCAGCTTGACAGAGTTGGAG 300  
DB 360 CAACTGTTGAAGAGCTATGAAATATGATTTGCTCTTTCAGCTTGACAGAGTTGGAG 419  
OY 301 TATGCAAAAGCTATATTTTGCAGAAAGAAATTAATCTCTCTGAAACATCTTAAAGAT 360  
DB 420 TATGCAAAAGCTATATTTTGCAGAAAGAAATTAATCTCTCTGAAACATCTTAAAGAT 479  
OY 361 GAAATTTCTATCATCAAACTATGAGGCTACAGAAAGCCGTCACAAAGACTTTCACAGAT 420  
DB 480 GAAATTTCTATCATCAAACTATGAGGCTACAGAAAGCCGTCACAAAGACTTTCACAGAT 539  
OY 421 GAACCCGAAATCTCTTGCAGAAAGAGCTGCTGATCAACCTCTTAACCTTGA 480  
DB 540 GAACCCGAAATCTCTTGCAGAAAGAGCTGCTGATCAACCTCTTAACCTTGA 599  
OY 481 ACTGTGAGAACTCTGAGAGCAAGAGAGCTACAACTCAAAAGAGCTGCTTACAT 540  
DB 600 ACTGTGAGAACTCTGAGAGCAAGAGAGCTACAACTCAAAAGAGCTGCTTACAT 659  
OY 541 GAATTTGGATCTGATTTCTGTAAGATACCGTTATATAGCAACTATTTGACAGTGGGA 600  
DB 660 GAATTTGGATCTGATTTCTGTAAGATACCGTTATATAGCAACTATTTGACAGTGGGA 719  
OY 601 GATCAAGATTTGTACAAATCACCCCTCAAGAAAGCGGATGAAATCAGTTGGATTCT 660  
DB 720 GATCAAGATTTGTACAAATCACCCCTCAAGAAAGCGGATGAAATCAGTTGGATTCT 779

OY 661 GCAAAAAAGCGTCTGTGATTTTCTGAGACGAGTAAACAATTAATGATCAATCA 720  
DB 780 GCAAAAAAGCGTCTGTGATTTTCTGAGACGAGTAAACAATTAATGATCAATCA 839  
OY 721 CCCAGTATATGATTTTGAACACCACTGAGAGCGTGCAGCTGAGAGCATCCAGAAAG 780  
DB 840 CCCAGTATATGATTTTGAACACCACTGAGAGCGTGCAGCTGAGAGCATCCAGAAAG 899  
OY 781 TATCAGGCTGATCTGTTTCAAACTTCAGTGTGAGCGCATGTCGACAAATATCACTGCC 840  
DB 900 TATCAGGCTGATCTGTTTCAAACTTCAGTGTGAGCGCATGTCGACAAATATCACTGCC 959  
OY 841 AGCTCATTAAGCATGAGAACAGAGCTTTATTAATCACTAAAGACAGATGATGAGAA 900  
DB 960 AGCTCATTAAGCATGAGAACAGAGCTTTATTAATCACTAAAGACAGATGATGAGAA 1019  
OY 901 AAGGCTGAATTTCTGTATTAAGCAAAAGCCTGCTTACAGAGAGCCCAATACAGA 960  
DB 1020 AAGGCTGAATTTCTGTATTAAGCAAAAGCCTGCTTACAGAGAGCCCAATACAGA 1079  
OY 961 TGGGCTGGAAGTAAAGAAACATGTAATGATGAGGCGGAGCTCCACAGAGAAAAAGGTA 1020  
DB 1080 TGGGCTGGAAGTAAAGAAACATGTAATGATGAGGCGGAGCTCCACAGAGAAAAAGGTA 1139  
OY 1021 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGCAAAACTGCCATGC 1080  
DB 1140 GATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGATGAAATAGCAAAACTGCCATGC 1199  
OY 1081 TCAGAGATCCCTAGAGATCTGAAGATGTTCTTGATTAACATTAATAGCAGATTACAG 1140  
DB 1200 TCAGAGATCCCTAGAGATCTGAAGATGTTCTTGATTAACATTAATAGCAGATTACAG 1259  
OY 1141 AAAATTAATGAGTGTCTTCCAGAGAGTGAATGATTTAGGTTGATGATGATCAATGAT 1200  
DB 1260 AAAATTAATGAGTGTCTTCCAGAGAGTGAATGATTTAGGTTGATGATGATCAATGAT 1319  
OY 1201 GGGGAGTCTGATCAATCAATGCGCAAAAGAGTGTATTTGACGCTTCAATAGAGTATGAT 1260  
DB 1320 GGGGAGTCTGATCAATCAATGCGCAAAAGAGTGTATTTGACGCTTCAATAGAGTATGAT 1379  
OY 1261 GAATATTTCTGTTCTTACAGAGAAATAGACTTACTGCGCAGTATCTCATAGGCTTTTA 1320  
DB 1380 GAATATTTCTGTTCTTACAGAGAAATAGACTTACTGCGCAGTATCTCATAGGCTTTTA 1439  
OY 1321 ATATGTAAGTGAAGAGTTCACCTCAATCACTAAGAGTATATTTGAAGCAAAATA 1380  
DB 1440 ATATGTAAGTGAAGAGTTCACCTCAATCACTAAGAGTATATTTGAAGCAAAATA 1499  
OY 1381 TTTGGGAAACCTATCGGAAGAGGCAAGGCTCCCAACTTAAGCCATGTAACAGAAAT 1440  
DB 1500 TTTGGGAAACCTATCGGAAGAGGCAAGGCTCCCAACTTAAGCCATGTAACAGAAAT 1559  
OY 1441 CTAAATTTAGAGCATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAAAT 1500  
DB 1560 CTAAATTTAGAGCATTTGTTACTGAGCCACAGATTAATACAGAGCGTCCCTCACAAAT 1619  
OY 1501 AAATTAAGGTAAGAGAGAGCTATACATACAGGCTTATCTCTGAGATTTTATCAAAA 1560  
DB 1620 AAATTAAGGTAAGAGAGAGCTATACATACAGGCTTATCTCTGAGATTTTATCAAAA 1679  
OY 1561 GCAGATTTGGCAGTTCAAAAGAGCTCTGAATGTAATACAGGAACTAAACCAAGCGAG 1620  
DB 1680 GCAGATTTGGCAGTTCAAAAGAGCTCTGAATGTAATACAGGAACTAAACCAAGCGAG 1739  
OY 1621 CAGATGCTCAAGTGAATGATATTACTAATATGCTGATGAGATTAACCAAAAGGTGAT 1680  
DB 1740 CAGATGCTCAAGTGAATGATATTACTAATATGCTGATGAGATTAACCAAAAGGTGAT 1799  
OY 1681 TCTATTAGAAATGGAAGAAATCCTAACCAATGAAATCAGTCCGAAAGAAATGCTGCTTC 1740  
DB 1800 TCTATTAGAAATGGAAGAAATCCTAACCAATGAAATCAGTCCGAAAGAAATGCTGCTTC 1859  
OY 1741 AAAACGAAGCTGAACCTATAGCAGCATATAAGCAATATGAACTGAAATTAATATC 1800



Db	1860	AAAAGCAAGGCTGACCTTTAAAGCAGCACTATAGCAATATGAACTCGAATTAATATC	1919
Qy	1801	CACAAATTCAAAAGCACCTAAAAAGATAGCTGAGAGGAAGTCTTACAGGCAATAT	1860
Db	1920	CACAATTCAAAAGCACCTAAAAAGAAATAGCGTAGAGAGAACTCTTACACAGCATAT	1979
Qy	1861	CATGGCCTTGAACCTAGTACTCAGTGAATCTTAAGCCCACTAATTGTACTGAATTCGA	1920
Db	1980	CATGGCCTTGAACCTAGTACTCAGTGAATCTTAAGCCCACTAATTGTACTGAATTCGA	2039
Qy	1921	ATTGATAGTGTGTTTAGCTAGCTAGAGGATAAAGAAAAAATAATACACCAATGGCACT	1980
Db	2040	ATTGATAGTGTGTTTCTTAGCTAGCTAGAGGATAAAGAAAAAATAATACACCAATGGCACT	2099
Qy	1981	AGGCACAGCAGAAACCTACACCTATGGAAGGTAAAGAACTGCAACTGAGGACCAAG	2040
Db	2100	AGGCACAGCAGAAACCTACACCTATGGAAGGTAAAGAACTGCAACTGAGGACCAAG	2159
Qy	2041	AGTACAAAGCCAAATGAACAGACAGTAAAGACATGACAGTACTTTCCAGAGCTG	2100
Db	2160	ATAACAACCCAAATGAACAGACAGTAAAGACATGACAGTACTTTCCAGAGCTG	2219
Qy	2101	AAGTTAACAAATGACACCTGGTCTTCTTCTAGTGTCAAAATACACAGTAACCTTAAGCA	2160
Db	2220	AAGTTAACAAATGACACCTGGTCTTCTTCTAGTGTCTTAATATACAGTGAACCTTAAGCA	2279
Qy	2161	TTTGTCAATCCTAGCCTTCCAAAGAGAGAAAAAGAGAAACTGAACAAGTTAAAGTG	2220
Db	2280	TTTGTCAATCCTAGCCTTCCAAAGAGAGAAAAAGAGAAACTGAACAAGTTAAAGTG	2339
Qy	2221	TCTAATTAATGCTGAAGAGACCCCAAGATCTCATGTTAACTGGAGAAAGGTTTCCAACT	2280
Db	2340	TCTAATTAATGCTGAAGAGACCCCAAGATCTCATGTTAACTGGAGAAAGGTTTCCAACT	2399
Qy	2281	GAAAGATCTGTAGAGAGTAGAGCATTTGTCATTTGTCACCTGGTACTGATTAAGCACTCAG	2340
Db	2400	GAAAGATCTGTAGAGAGTAGAGCATTTGTCATTTGTCACCTGGTACTGATTAAGCACTCAG	2459
Qy	2341	GAAAGATCTGTTACTGGAAGTTTAGCACTAGGGAAGGCAAAAACGAACCAATTA	2400
Db	2460	GAAAGATCTGTTACTGGAAGTTTAGCACTAGGGAAGGCAAAAACGAACCAATTA	2519
Qy	2401	TGTGTGAGTCAAGTGTGCGACATTTTGAAACCCCAAGGAGCTAATTCATGTTGTTCCAA	2460
Db	2520	TGTGTGAGTCAAGTGTGCGACATTTTGAAACCCCAAGGAGCTAATTCATGTTGTTCCAA	2579
Qy	2461	GATATTAAGAAATGACACAGAAAGGCTTAAATGATCCATTGGGACAAATGAATTAACACAGT	2520
Db	2580	GATATTAAGAAATGACACAGAAAGGCTTAAATGATCCATTGGGACAAATGAATTAACACAGT	2639
Qy	2521	CGGGAACAGCATAGAAATGGAAGAAAGTGAATGCTCAGTATTTGSCAATATCA	2580
Db	2640	CGGGAACAGCATAGAAATGGAAGAAAGTGAATGCTCAGTATTTGSCAATATCA	2699
Qy	2581	TTTCAAGGTTTCAAAAGCCCACTCATTTGCTGTGTTTCAAAATCCAGAAATGCAAGAG	2640
Db	2700	TTTCAAGGTTTCAAAAGCCCACTCATTTGCTGTGTTTCAAAATCCAGAAATGCAAGAG	2759
Qy	2641	GAAATGCAACATTTCTGGCCCACTCTGGGCTTTAAAGAAACAAAGTCCAAAAGTCACT	2700
Db	2760	GAAATGCAACATTTCTGGCCCACTCTGGGCTTTAAAGAAACAAAGTCCAAAAGTCACT	2819
Qy	2701	TTTGAATGTGAACAAAAGGAAGAAATCAAGAAAGATGCTAATATACAGCCCTGA	2760
Db	2820	TTTGAATGTGAACAAAAGGAAGAAATCAAGAAAGATGCTAATATACAGCCCTGA	2879
Qy	2761	CAGACAGTTAATATCACTGACAGGCTTTCCTGGTGTGGTGAAGAAAGATTAAGCCAGTTGAT	2820
Db	2880	CAGACAGTTAATATCACTGACAGGCTTTCCTGGTGTGGTGAAGAAAGATTAAGCCAGTTGAT	2939
Qy	2821	AATGCCAAATGTAGATCAAGAGAGGCTCTAGTTTTGTATATCATCTCAGTTTCAGAGC	2880

Db	2940	AATGCCAAATGTATGATATCAAGAGAGGCTCTAGGTTTGTCTATCATCTCAGTTTCAGAGGC	2991
QY	2881	AAGCAACTGAGACTCTTACTCCAAATTAACATGAGACTTTACAAAAACCCATATCGTATA	2944
Db	3000	AAGCAACTGAGACTCTTACTCCAAATTAACATGAGACTTTACAAAAACCCATATCGTATA	3051
QY	2941	CCACGACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAATCTGCTAGAG	3000
Db	3060	CCACGACTTTTCCATCAAGTCAATTTGTTAAACTAAATGTAAGAAAATCTGCTAGAG	3111
QY	3001	GAAACTTTGAGGAACATTCATGATGACCTGGAAAGGAATGGAAATGAGAAACATTTCCA	3066
Db	3120	GAAACTTTGAGGAACATTCATGATGACCTGGAAAGGAATGGAAATGAGAAACATTTCCA	3177
QY	3061	AGTACAGTGGAGCACTTACCGGTATATACATTAGAGAAAATGTTTTTAAACACACGAC	3122
Db	3180	AGTACAGTGGAGCACTTACCGGTATATACATTAGAGAAAATGTTTTTAAAGACACGAC	3233
QY	3121	TCAAGCAATATTAAATGAGTTCAGTCTACTAATGAAGTGGCTCCAGTATTAATGAA	3180
Db	3240	TCAAGCAATATTAAATGAGTTCAGTCTACTAATGAAGTGGCTCCAGTATTAATGAA	3299
QY	3181	ATAGGTTCCAGATGAAACATTCAGAGAGAACTAGTAAAGACAGAGGCCCCAAATTTG	3244
Db	3300	ATAGGTTCCAGATGAAACATTCAGAGAGAACTAGTAAAGACAGAGGCCCCAAATTTG	3355
QY	3241	AATGCTATGCTTAGATTAGGGGTTTGGCACTGAGAGTCTATAACAAAGTTCTCTGGA	3300
Db	3360	AATGCTATGCTTAGATTAGGGGTTTGGCACTGAGAGTCTATAACAAAGTTCTCTGGA	3419
QY	3301	AGTAATTTGTAAGCAATCTGAAATTAAGAAAGCAAGATATGAAAGATAGTTCAGACTGT	3366
Db	3420	AGTAATTTGTAAGCAATCTGAAATTAAGAAAGCAAGATATGAAAGATAGTTCAGACTGT	3479
QY	3361	AATACAGATTTCTCTCCATATCTGATTTTCAGATTAAGAACAGCTATGGAGATAGT	3420
Db	3480	AATACAGATTTCTCTCCATATCTGATTTTCAGATTAAGAACAGCTATGGAGATAGT	3539
QY	3421	CATGCATCTCAGGTTGTTCTGAGACACCTGATGACTGTTAGATGATGGTGAATAAG	3480
Db	3540	CATGCATCTCAGGTTGTTCTGAGACACCTGATGACTGTTAGATGATGGTGAATAAG	3599
QY	3481	GAAGATCTAGTTTGTGTAANAATGACATTAAGGAAAGTCTGCTGTTTTAGCAAAAGC	3540
Db	3600	GAAGATCTAGTTTGTGTAANAATGACATTAAGGAAAGTCTGCTGTTTTAGCAAAAGC	3659
QY	3541	GTCCAGAAAGGAGACTTACGACGAGTCTGACCTTTCACCCATACACATTTGGCTAG	3600
Db	3660	GTCCAGAAAGGAGACTTACGACGAGTCTGACCTTTCACCCATACACATTTGGCTAG	3719
QY	3601	GATTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTTATCTAGTAGGAT	3660
Db	3720	GATTACCGAAGAGGGGCCAAGAAATTAAGTCTCAGAGAGAACTTATCTAGTAGGAT	3779
QY	3661	GAAGAGCTTCCCTGCTTCCAACTTTGTTATTTGGTAAAGTAACAAATATACCTTCTCAG	3720
Db	3780	GAAGAGCTTCCCTGCTTCCAACTTTGTTATTTGGTAAAGTAACAAATATACCTTCTCAG	3839
QY	3721	TCTACTAGGCATAGACCGCTGTCTACGAGTGTCTGTTAAGAACACAGAGAGAAATTTA	3780
Db	3840	TCTACTAGGCATAGACCGCTGTCTACGAGTGTCTGTTAAGAACACAGAGAGAAATTTA	3899
QY	3781	TTATCATTTGAAGAAATAGCTTAAATGACATCGCAATCCAGTAATATTTGGCAAGGCAATCT	3840
Db	3900	TTATCATTTGAAGAAATAGCTTAAATGACATCGCAATCCAGTAATATTTGGCAAGGCAATCT	3959
QY	3841	CAGGAACATCAGCTTAGTAGAGAGAAACAAATGTTCTGCTACTGTTTCTTCACAGTGC	3900
Db	3960	CAGGAACATCAGCTTAGTAGAGAGAAACAAATGTTCTGCTACTGTTTCTTCACAGTGC	4019
QY	3901	AGTGAATTTGAAGACTTGACTGCAATATCAAAACCCAGATCTCTTCTTGATTTGCTTCT	3960
Db	4020	AGTGAATTTGAAGACTTGACTGCAATATCAAAACCCAGATCTCTTCTTGATTTGCTTCT	4079



QY 3961 TCACAAATGAGGATCAGTCTGAAAGCCAGGAGTGTGTGATGACAGGAATG 4020  
 DB 4080 TCACAAATGAGGATCAGTCTGAAAGCCAGGAGTGTGTGATGACAGGAATG 4139  
 QY 4021 GTTTCATATGAGGAAGAGGAGCGGCTTGGAAAGAAATATCAAGAGCAAGC 4080  
 DB 4140 GTTTCATATGAGGAAGAGGAGCGGCTTGGAAAGAAATATCAAGAGCAAGC 4199  
 QY 4081 ATGATTCATTAAGTGTGAGAGCAGATCTGGGTGTGAGAGTGAACAGAGTCTGAA 4140  
 DB 4200 ATGATTCATTAAGTGTGAGAGCAGATCTGGGTGTGAGAGTGAACAGAGTCTGAA 4259  
 QY 4141 GACTGTACAGGCTATCTCTCAGAGTACATTTTAAACATCAGCAGAGGATACATG 4200  
 DB 4260 GACTGTACAGGCTATCTCTCAGAGTACATTTTAAACATCAGCAGAGGATACATG 4319  
 QY 4201 CAACATTAACCTGAATGAAGTCCAGAGGAATGGCTGATAGAGTGTGTGAACAG 4280  
 DB 4320 CAACATTAACCTGAATGAAGTCCAGAGGAATGGCTGATAGAGTGTGTGAACAG 4379  
 QY 4261 CATGGAGCAGAGCTTCTTAACAGTACCTTCCATCAATAGTACTCTTCCCTTGAG 4320  
 DB 4380 CATGGAGCAGAGCTTCTTAACAGTACCTTCCATCAATAGTACTCTTCCCTTGAG 4439  
 QY 4321 GACCTGCAAAATCCAGAAAGCAGATCAGAAAAAGCAGTATTAACCTCAGAAAAAGT 4380  
 DB 4440 GACCTGCAAAATCCAGAAAGCAGATCAGAAAAAGCAGTATTAACCTCAGAAAAAGT 4499  
 QY 4381 AGTGAATACCTATTAAGCAGATCCAGAAAGCCTTCTGTGCAAGTGTGAGTCT 4440  
 DB 4500 AGTGAATACCTATTAAGCAGATCCAGAAAGCCTTCTGTGCAAGTGTGAGTCT 4559  
 QY 4441 GCAGATAGTCTTACCACTAAATTAAGAAACAGAGAGTGAAGAGTATCCCTTCTAAA 4500  
 DB 4560 GCAGATAGTCTTACCACTAAATTAAGAAACAGAGAGTGAAGAGTATCCCTTCTAAA 4619  
 QY 4501 TGCCCATCATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
 DB 4620 TGCCCATCATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4679  
 QY 4561 AACACCATCTCAAGAGAGCTCATTAAGTGTGTGATGATGATGATGATGATGATGATGAT 4620  
 DB 4680 AACACCATCTCAAGAGAGCTCATTAAGTGTGTGATGATGATGATGATGATGATGATGAT 4739  
 QY 4621 GAGTCTGGCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4680  
 DB 4740 GAGTCTGGCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4799  
 QY 4681 ACCCCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
 DB 4800 ACCCCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4859  
 QY 4741 GAAGAGAGAGCCAGAGTACCTGCTGTGGAACATACCATCTTCACTGATGATGATGATGAT 4800  
 DB 4860 GAAGAGAGAGCCAGAGTACCTGCTGTGGAACATACCATCTTCACTGATGATGATGATGAT 4919  
 QY 4801 AAAGTTCATTAAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4860  
 DB 4920 AAAGTTCATTAAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4979  
 QY 4861 GATACGTCTGGATTAATGCAATGAGAAAGTGTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAT 4920  
 DB 4980 GATACGTCTGGATTAATGCAATGAGAAAGTGTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAT 5039  
 QY 4921 GCTTCAACAGAAAGGCTCAACAAAGATGTCCATGATGATGATGATGATGATGATGATGATGAT 4980  
 DB 5040 GCTTCAACAGAAAGGCTCAACAAAGATGTCCATGATGATGATGATGATGATGATGATGATGAT 5099  
 QY 4981 GAATTTATGCTGTGTCAAGTGTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAT 5040  
 DB 5100 GAATTTATGCTGTGTCAAGTGTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAT 5159

QY 5041 ACTGAGAGACTACTATGTTGTTATGAAGACAGATCTGAGTTGTGTGAACGAGACA 5100  
 DB 5160 ACTGAGAGACTACTATGTTGTTATGAAGACAGATCTGAGTTGTGTGAACGAGACA 5219  
 QY 5101 CTGAAATATTTCTAGGAATTCGGGAGGAAATGGGTACTGATGATGATGATGATGATGATGAT 5160  
 DB 5220 CTGAAATATTTCTAGGAATTCGGGAGGAAATGGGTACTGATGATGATGATGATGATGATGAT 5279  
 QY 5161 CAGTCTATTAAGAAAGAAATGCTGAATGAGATGATGATGATGATGATGATGATGATGATGATGAT 5220  
 DB 5280 CAGTCTATTAAGAAAGAAATGCTGAATGAGATGATGATGATGATGATGATGATGATGATGATGAT 5339  
 QY 5221 GTCATGAGAGAAACCAAGAGTCCAAAGCAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGATC 5280  
 DB 5340 GTCATGAGAGAAACCAAGAGTCCAAAGCAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGATC 5399  
 QY 5281 TTGAGGGGCTAGAAATCTGTGTATGGGCTTCCAAACATGCTCCAGAGATCAACTG 5340  
 DB 5400 TTGAGGGGCTAGAAATCTGTGTATGGGCTTCCAAACATGCTCCAGAGATCAACTG 5459  
 QY 5341 GAATGAT 5400  
 DB 5460 GAATGAT 5519  
 QY 5401 GGCACAGGTGTCCACCAATTTGTGTTGTGACAGCAGATGATGATGATGATGATGATGATGATGATGAT 5460  
 DB 5520 GGCACAGGTGTCCACCAATTTGTGTTGTGACAGCAGATGATGATGATGATGATGATGATGATGATGAT 5579  
 QY 5461 TTCCATGAT 5520  
 DB 5580 TTCCATGAT 5639  
 QY 5521 AGTGTACACTCTTACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 5580  
 DB 5640 AGTGTACACTCTTACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 5699  
 QY 5581 AGCCACTAC 5589  
 DB 5700 AGCCACTAC 5708

RESULT 5  
 US-10-022-819-1  
 : Sequence 1, Application US/10022819  
 : Publication No. US20030027166A1  
 : GENERAL INFORMATION:  
 : APPLICANT: ALLEN, Antonette C. P.  
 : OLSEN, Sheri J.  
 : LAWRENCE, Tammy  
 : ANGELLY, Tracy S.  
 : RABIN, Mark B.  
 : TITLE OF INVENTION: CODING SEQUENCE HAPLOTYPE OF THE HUMAN  
 : BRCA1 GENE  
 : NUMBER OF SEQUENCES: 67  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Morgan Lewis & Bockius LLP  
 : STREET: 1111 Pennsylvania Avenue  
 : CITY: Washington DC  
 : STATE: District of Columbia  
 : COUNTRY: USA  
 : ZIP: 20004  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent In Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/10/022,819  
 : FILING DATE: 22-Apr-2002  
 : CLASSIFICATION: <Unknown>  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 09/074,452  
 : FILING DATE: 1998-05-06

```

ATTORNEY/AGENT INFORMATION:
NAME: <unknown>
REGISTRATION NUMBER: <unknown>
REFERENCE/DOCKET NUMBER: 044921-5049-01-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-739-3000
TELEFAX: 202-739-3001
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5711 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
STRAIN: BRCA1
HAPLOTYPE: OM14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-022-819-1

```

Query Match	99.98;	Score 5582.6;	DB 9;	Length 5711;
-------------	--------	---------------	-------	--------------

```

best local similarity: 99.98; freq: 0;
Matches 5585; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY	1	TTGGATTATATCGCTTCGCGGTTGAACAAAGTACAAAAATGTCATTAATGCTATGCGAGAA	60
Db	120	ATGAGATTATTCGCTTCGCGGTTGAACAAAGTACAAAAATGTCATTAATGCTATGCGAGAA	179
OY	61	ATCTTAGAGTCCCATCTGCTCTGAGTGGATCAAGAACCTGTCTCCACAAAGTGTAC	120
Db	180	ATCTTAGAGTCCCATCTGCTCTGAGTGGATCAAGAACCTGTCTCCACAAAGTGTAC	239
OY	121	CACATATTTCGCAAAATTTTGCATGCTGAAACTTCACACCGAAGAAAGGGCCTTCACAG	180
Db	240	CACATATTTCGCAAAATTTTGCATGCTGAAACTTCACACCGAAGAAAGGGCCTTCACAG	299
OY	181	TGTCCTTTATGTAAAGAAATGATATAACCAAAAGAGCCTACAAAGAAAGTACGAGATTAGT	240
Db	300	TGTCCTTTATGTAAAGAAATGATATAACCAAAAGAGCCTACAAAGAAAGTACGAGATTAGT	359
OY	241	CACCTGTGTAAGAGCTATGTAATAATCTTTGTGCTTTGCAGCTGACACAGGTTGGAG	300
Db	360	CACCTGTGTAAGAGCTATGTAATAATCTTTGTGCTTTGCAGCTGACACAGGTTGGAG	419
OY	301	TATGCAAAACAGCTATTAATTTTGCAAAAAAGGAAATAACTCTCCTGGAACATCTAAAGAT	360
Db	420	TATGCAAAACAGCTATTAATTTTGCAAAAAAGGAAATAACTCTCCTGGAACATCTAAAGAT	479
OY	361	GAACTTCATCATCCAAAGTATGGGCTACAGAAACCGTGGCCAAAGACTCTACAGAT	420
Db	480	GAACTTCATCATCCAAAGTATGGGCTACAGAAACCGTGGCCAAAGACTCTACAGAT	539
OY	421	GAACCCGAAATCCTTCCTTCAGAGAAACCAAGTCTCACTGTCCAACTCTTAACCTTGA	480
Db	540	GAACCCGAAATCCTTCCTTCAGAGAAACCAAGTCTCACTGTCCAACTCTTAACCTTGA	599
OY	481	ACTGTGAGAACTCTGAGAGCAAAAGCAGCGGATACAACTCAAAGAGCTGTCTACTT	540
Db	600	ACTGTGAGAACTCTGAGAGCAAAAGCAGCGGATACAACTCAAAGAGCTGTCTACTT	659
OY	541	GAATTGGATCTGATTCTTCTGAAGATACCGCTTAATAGGCAACTTATTCGAGTGGGA	600
Db	660	GAATTGGATCTGATTCTTCTGAAGATACCGCTTAATAGGCAACTTATTCGAGTGGGA	719
OY	601	GATCAGATTTGTTCAAAATCACCCCTCAGGAACCAAGGATGAATCAGTTGGATTCT	660

Db	720	GATCAGAAATTTGTTACAAATCACCCCTCCAGGAACACGAGGATGAAATCAGTTTGATTTCT	779
QY	661	GCAAAAAAGCGCTGCTTGAAATTTTCTGAGACGGATGTACAAATPACTGAACATCATCA	720
Db	780	GCAAAAAAGCGCTGCTTGAAATTTTCTGAGACGGATGTACAAATPACTGAACATCATCA	839
QY	721	CCGAGTAAATGATTTGACACCACTGAGAGGGTGCAGCTGAGAGGCATCCGAAAG	780
Db	840	CCGAGTAAATGATTTGACACCACTGAGAGGGTGCAGCTGAGAGGCATCCGAAAG	899
QY	781	TATCAGGCTAGTTCCTGTTTCAAACTTGCATGTGAGCCATGTGGACAAATPACTCATGCC	840
Db	900	TATCAGGCTAGTTCCTGTTTCAAACTTGCATGTGAGCCATGTGGACAAATPACTCATGCC	959
QY	841	AGCTCATTTACAGCATGAGACAGCATGTTTATTTACTCACTAAAGACAGATGATGTAGAA	900
Db	960	AGCTCATTTACAGCATGAGACAGCATGTTTATTTACTCACTAAAGACAGATGATGTAGAA	1019
QY	901	AAGCGTGAATTCCTGTAATTAAGCAACAGCGCTGGCTAGCAAGAGCCAACTATACAGA	960
Db	1020	AAGCGTGAATTCCTGTAATTAAGCAACAGCGCTGGCTAGCAAGAGCCAACTATACAGA	1079
QY	961	TGGCGTGAAGTAAAGGAACAATGTTATGATAGGCGGACATCCACACAGAAAAAGGTA	1020
Db	1080	TGGCGTGAAGTAAAGGAACAATGTTATGATAGGCGGACATCCACACAGAAAAAGGTA	1139
QY	1021	GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAAGCAGAAACTCCATGC	1080
Db	1140	GATCTGAATGCTGATCCCTGTGTGAGAGAAAGATGGAATTAAGCAGAAACTCCATGC	1199
QY	1081	TCAGGAATCCTAGAGTACTGAAGAGTGTCCCTGGATACACATAATAGCAGCATTCAG	1140
Db	1200	TCAGGAATCCTAGAGTACTGAAGAGTGTCCCTGGATACACATAATAGCAGCATTCAG	1259
QY	1141	AAAGTAAATGAGTGGTTTTCCAGAAAGTATGAACCTGTAGGTTTCGATGATCACAATGAT	1200
Db	1260	AAAGTAAATGAGTGGTTTTCCAGAAAGTATGAACCTGTAGGTTTCGATGATCACAATGAT	1319
QY	1201	GGGAGTCGTGAATCAAAATGCCAAAGTAGCTGATGTTTGAACGTTCTTAATAGAGTACAT	1260
Db	1320	GGGAGTCGTGAATCAAAATGCCAAAGTAGCTGATGTTTGAACGTTCTTAATAGAGTACAT	1379
QY	1261	GAATATTTCTGGTCTTTCAGAGAAATAGACTTACTGGCAGTGCATCATGAGGCTTTA	1320
Db	1380	GAATATTTCTGGTCTTTCAGAGAAATAGACTTACTGGCAGTGCATCATGAGGCTTTA	1439
QY	1321	ATATGTAAAGGGAAGAGCTTCACGCCAAATCAGTAGAGAGTAAATATGAGACAAATA	1380
Db	1440	ATATGTAAAGGGAAGAGCTTCACGCCAAATCAGTAGAGAGTAAATATGAGACAAATA	1499
QY	1381	TTTGGGAAAACCTATCGGAGAGGCAAGCCTCCCAACTTAAGCCATGTAACTGAAAT	1440
Db	1500	TTTGGGAAAACCTATCGGAGAGGCAAGCCTCCCAACTTAAGCCATGTAACTGAAAT	1559
QY	1441	CTAATTTATAGAGCAATTTGTTACTGAGCACACAGATAATACAAGAGCTGCCCTACAAAT	1500
Db	1560	CTAATTTATAGAGCAATTTGTTACTGAGCACACAGATAATACAAGAGCGCCCTACAAAT	1619
QY	1501	AAATTTAAAGCGTAAAGAGAGACCTACATCAGGCCCTCATCCTGTGGAGTTTATACAAA	1560
Db	1620	AAATTTAAAGCGTAAAGAGAGACCTACATCAGGCCCTCATCCTGTGGAGTTTATACAAA	1679
QY	1561	GCAGATTTGGCAGTTCCAAAGACCTCCTGAATATGATAATCAGGAGACTTAACCAACGGAG	1620
Db	1680	GCAGATTTGGCAGTTCCAAAGACCTCCTGAATATGATAATCAGGAGACTTAACCAACGGAG	1739
QY	1621	CAGAAATGGTCAAGTGATCATATTTACTATATAGTGTGATGAGANTAAACAAAAGTGAT	1680
Db	1740	CAGAAATGGTCAAGTGATCATATTTACTATATAGTGTGATGAGANTAAACAAAAGTGAT	1799
QY	1681	TCTATTTGAGATGAGAAAAATCTTAACCCATTAACATCACCTCGAAAAAGATCTGCTTTC	1740
Db	1800	TCTATTTGAGATGAGAAAAATCTTAACCCATTAACATCACCTCGAAAAAGATCTGCTTTC	1859

1741 AAAAGAAAGCTGAACCTATAGCAGCAGTATAAGCAATATGGAATCGAATTAATATC 1800  
1860 AAAAGAAAGCTGAACCTATAGCAGCAGTATAAGCAATATGGAATCGAATTAATATC 1919  
1801 CACAAATTCAAAAGCAGCTAAAGAAATAGGCTGAGGAGAAAGTCTTCTACCGAGCATAT 1860  
1920 CACAAATTCAAAAGCAGCTAAAGAAATAGGCTGAGGAGAAAGTCTTCTACCGAGCATAT 1979  
1861 CATGCGCTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1920  
1980 CATGCGCTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2039  
1921 ATTGATATGTTGTTTACAGTGAAGAGATTAAGAAAAAAGTTCACCAATATCCAGTC 1980  
2040 ATTGATATGTTGTTTACAGTGAAGAGATTAAGAAAAAAGTTCACCAATATCCAGTC 2039  
1981 AGGCACAGCAGAAACCTCAACATCATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2040  
2100 AGGCACAGCAGAAACCTCAACATCATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2159  
2041 AGTACAGCAAGCAATGAAGAGACAGTAAAGACATGACATGATCTTCTCCAGACTG 2100  
2160 AGTACAGCAAGCAATGAAGAGACAGTAAAGACATGACATGATCTTCTCCAGACTG 2219  
2101 AAGTTACAAATGACAGCTGTTCTTTACTAGTGTCAATATACAGTGAAGTGAAGTGAAGTGAAG 2160  
2220 AAGTTACAAATGACAGCTGTTCTTTACTAGTGTCAATATACAGTGAAGTGAAGTGAAGTGAAG 2279  
2161 TTTGTCAATCCCTGACCTTCAAGAGAAAGAAAGAAAGAAAGAAAGAAAGTGAAGTGAAGTGAAG 2220  
2280 TTTGTCAATCCCTGACCTTCAAGAGAAAGAAAGAAAGAAAGAAAGTGAAGTGAAGTGAAGTGAAG 2339  
2221 TCTAATATGCTGAGAGCCCAAGATCTCATGTTAGTGAAGAGAGGTTTTCAGACT 2280  
2340 TCTAATATGCTGAGAGCCCAAGATCTCATGTTAGTGAAGAGAGGTTTTCAGACT 2399  
2281 GAAAGATCTGTAGAGATGACAGTATTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2340  
2400 GAAAGATCTGTAGAGATGACAGTATTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2459  
2341 GAAAGATCTGTAGAGATGACAGTATTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2400  
2460 GAAAGATCTGTAGAGATGACAGTATTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2519  
2401 TGTGTGAT 2460  
2520 TGTGTGAT 2579  
2461 GATTAATGAATGACACAGAGGCTTTAATTCATTTGGACATGAAATTAACACAGT 2520  
2580 GATTAATGAATGACACAGAGGCTTTAATTCATTTGGACATGAAATTAACACAGT 2639  
2521 CGGGAACAAGCATAGAAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580  
2640 CGGGAACAAGCATAGAAATGGAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2699  
2581 TTTAAGGTTTCAAGGCGGCTGATTTGCTGTTTCAATTCAGAAATGCAAGAG 2640  
2700 TTTAAGGTTTCAAGGCGGCTGATTTGCTGTTTCAATTCAGAAATGCAAGAG 2759  
2641 GAAATGTCACATTTCTGCGCACTGCGGTCTTTAAGAAACAAAGTCCAAAGTCACT 2700  
2760 GAAATGTCACATTTCTGCGCACTGCGGTCTTTAAGAAACAAAGTCCAAAGTCACT 2819  
2701 TTTGATGTCACAAAGAAAGAAATCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760  
2820 TTTGATGTCACAAAGAAAGAAATCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2879  
2761 CAGACGTTAATATCACTGAGGCTTTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2820  
2880 CAGACGTTAATATCACTGAGGCTTTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2939

2821 AATGCCAATGTTAGTATCAAGAGAGGCTAGGTTTGTCTATCATCTCACTGACAGGC 2880  
2940 AATGCCAATGTTAGTATCAAGAGAGGCTAGGTTTGTCTATCATCTCACTGACAGGC 2999  
2881 AACGAACCTGACATCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATGATTA 2940  
3000 AACGAACCTGACATCTTACTCCAAATTAACATGAGCTTTTACAAAACCATATGATTA 3059  
2941 CCACCACTTTTCCCATCAAGTCAATTTGTTTAAACTAAATGTAAGAAAAATCTCTAGAG 3000  
3060 CCACCACTTTTCCCATCAAGTCAATTTGTTTAAACTAAATGTAAGAAAAATCTCTAGAG 3119  
3001 GAAACCTTTGAGAAACATTCATATGTCACCTGCAAGAGAAATGGAATGAGAACATTCGA 3060  
3120 GAAACCTTTGAGAAACATTCATATGTCACCTGCAAGAGAAATGGAATGAGAACATTCGA 3179  
3061 AGTACAGTGAAGCACAATTTAGCCGTTAATTAAGTGAAGAAATGTTTTTAAAGAACCCAGC 3120  
3180 AGTACAGTGAAGCACAATTTAGCCGTTAATTAAGTGAAGAAATGTTTTTAAAGAACCCAGC 3239  
3121 TCAAGCAATTTAATGAAGTGAAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3180  
3240 TCAAGCAATTTAATGAAGTGAAGTTCAGTACTAATGAAGTGGCTCCAGTATTAATGA 3299  
3181 ATAGGTTCCAGTGAAGAAACATTCAGAGAACTAGTGAAGAAAGAGGCCCCAAATTTG 3240  
3300 ATAGGTTCCAGTGAAGAAACATTCAGAGAACTAGTGAAGAAAGAGGCCCCAAATTTG 3359  
3241 AATGCTATGCTTTAGATTAGGGGTTTTGCCAACCTGAGGCTTAATAACAAAGTCTTCTGGA 3300  
3360 AATGCTATGCTTTAGATTAGGGGTTTTGCCAACCTGAGGCTTAATAACAAAGTCTTCTGGA 3419  
3301 AGTAAATGTAAGCAATCCGTAATAAAGCAAAATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3360  
3420 AGTAAATGTAAGCAATCCGTAATAAAGCAAAATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3479  
3361 AATACAGATTTCTCTCATATCTGATTTCAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3420  
3480 AATACAGATTTCTCTCATATCTGATTTCAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3539  
3421 CATGATCTCAGGTTTGTTCGAGACACCTGATGACCTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAG 3480  
3540 CATGATCTCAGGTTTGTTCGAGACACCTGATGACCTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAG 3599  
3481 GAAGATCTGATGTTTGTGTAAGAAATGACATTAAGAAAGTGTCTGTTTTCAGAAAGC 3540  
3600 GAAGATCTGATGTTTGTGTAAGAAATGACATTAAGAAAGTGTCTGTTTTCAGAAAGC 3659  
3541 GTCCAGAAAGAGAGCTTTACAGAGTCTAGCCCTTTCACCCATACATTTTGGCTCAG 3600  
3660 GTCCAGAGAGAGAGCTTTACAGAGTCTAGCCCTTTCACCCATACATTTTGGCTCAG 3719  
3601 GGTTCAGAGAGAGAGCTTTACAGAGTCTAGCCCTTTCACCCATACATTTTGGCTCAG 3660  
3720 GGTTCAGAGAGAGAGCTTTACAGAGTCTAGCCCTTTCACCCATACATTTTGGCTCAG 3779  
3661 GAAGAGCTTCCCTGCTTCCCAACACTTGTATTTGTTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3720  
3780 GAAGAGCTTCCCTGCTTCCCAACACTTGTATTTGTTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3839  
3721 TCTACTAGCATAGCAGCTTGTGCTAGCAGTGTCTGCTAAGAAACAGAGAGAGATTTA 3780  
3840 TCTACTAGCATAGCAGCTTGTGCTAGCAGTGTCTGCTAAGAAACAGAGAGAGATTTA 3839  
3781 TTTATCATTTGAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3840  
3900 TTTATCATTTGAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3959  
3841 CAGGAACATACCTTATGAGAGAAACAAATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3900  
3960 CAGGAACATACCTTATGAGAGAAACAAATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4019  
3901 AGTGAATTTGAAGACTTGAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3960

```

|||||
4020 AGGAAATGGAAACCTTGATGCAAAATACAAACACCAGATCCTTTCTGATGGTCT 4079
|||
3961 TCCAAACAATGAGCATCATCTGAAAGCCAGAGATTTGCTGAGTGAAGAATG 4020
|||
4080 TCCAAACAATGAGCATCATCTGAAAGCCAGAGATTTGCTGAGTGAAGAATG 4139
|||
4021 GTTTCAGATGATGAAGAAGAGACGGCTTGGAAATAATATCAAGAAGCAAAAGC 4080
|||||
4140 GTTTCAGATGATGAAGAAGAGACGGCTTGGAAATAATATCAAGAAGCAAAAGC 4199
|||||
4081 ATGGATTCAACTTAGTGAAGAGATCTGGGTGTGAGAGTGAACAACGCTCTGAA 4140
|||||
4200 ATGGATTCAACTTAGTGAAGAGATCTGGGTGTGAGAGTGAACAACGCTCTGAA 4259
|||||
4141 GACTGTCAGAGGCTATCTCTCTCAGAGTGAATTTTAACTCAGCAGAGAGATACATG 4200
|||||
4260 GACTGTCAGAGGCTATCTCTCTCAGAGTGAATTTTAACTCAGCAGAGAGATACATG 4319
|||||
4201 CAACATACCTGTATAAGCTCCAGCAGAGAAATGGCTGAATAGAAAGCTGTAGAACAG 4260
|||||
4320 CAACATACCTGTATAAGCTCCAGCAGAGAAATGGCTGAATAGAAAGCTGTAGAACAG 4379
|||||
4261 CATGGAGAGCAGCTTCTTAACACCTTCCATCATTAAGTGAAGTCTTGTGCTTGAAG 4320
|||||
4380 CATGGAGAGCAGCTTCTTAACACCTTCCATCATTAAGTGAAGTCTTGTGCTTGAAG 4439
|||||
4321 GACCTCGAAATCCAGAACACATCATGAAAAAGCAATTTAACTTACAGAAAGT 4380
|||||
4440 GACCTCGAAATCCAGAACACATCATGAAAAAGCAATTTAACTTACAGAAAGT 4499
|||||
4381 AGTGAATACCTGTATAAGCTCCAGCAGAGAAATGGCTGTAGAACAGTGTGAGTGTCT 4440
|||||
4500 AGTGAATACCTGTATAAGCTCCAGCAGAGAAATGGCTGTAGAACAGTGTGAGTGTCT 4559
|||||
4441 GGAGATAGTCTTCCAGTAAATTAAGAAGCAGAGATGGAAGTGTATCCCTTCTGAAA 4500
|||||
4560 GGAGATAGTCTTCCAGTAAATTAAGAAGCAGAGATGGAAGTGTATCCCTTCTGAAA 4619
|||||
4501 TGCCCATCATTTAGATAGTGTGATGATGACAGATTTGCTGGAGTCTTCAAGAAATGA 4560
|||||
4620 TGCCCATCATTTAGATAGTGTGATGATGACAGATTTGCTGGAGTCTTCAAGAAATGA 4679
|||||
4561 AACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGCAACAGCTGAAA 4620
|||||
4680 AACTACCATCTCAAGAGAGCTCATTAAGTGTGATGATGAGAGCAACAGCTGAAA 4739
|||||
4621 GAGTCTGGGCGCACAGATTTGACGGAACATCTTACTTGCAGAGCAAGTCTTAGAGGGA 4680
|||||
4740 GAGTCTGGGCGCACAGATTTGACGGAACATCTTACTTGCAGAGCAAGTCTTAGAGGGA 4799
|||||
4681 ACCCCTTACCTGGAATCTGGAATCAGCTCTTCTGTGATGACCTGATGATCTCTCT 4740
|||||
4800 ACCCCTTACCTGGAATCTGGAATCAGCTCTTCTGTGATGACCTGATGATCTCTCTCT 4859
|||||
4741 GAAGACAGAGCCCGACAGCTCAGCTGTGTTGGCAACATACCATCTTCAACCTCTGCAATG 4800
|||||
4860 GAAGACAGAGCCCGACAGCTCAGCTGTGTTGGCAACATACCATCTTCAACCTCTGCAATG 4919
|||||
4801 AAAAGTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCCAGCTGTCTCATCTACT 4860
|||||
4920 AAAAGTCCCAATTTGAAGTGTGAGAAATCTGCCAGAGTCCAGCTGTCTCATCTACT 4979
|||||
4861 GATACGTGGGTATATGATGATGAAAGAGTGTGAGAGAGGAGAGAGAGATGACA 4920
|||||
4980 GATACGTGGGTATATGATGATGAAAGAGTGTGAGAGAGGAGAGAGAGATGACA 5039
|||||
4921 GCTTCAACAGAAAGGCTCAACAAAGATGTCCATGTTGCTGTGCTGACCCAGAA 4980
|||||
5040 GCTTCAACAGAAAGGCTCAACAAAGATGTCCATGTTGCTGTGCTGACCCAGAA 5099
|||||
4981 GAATTTATCTGCTGTACAGATTTGCGAGAAAACACACATCTTAACTAATCTAAT 5040
|||||

```

```

Db 5100 GAATTTATGCTGTGTACAGATTTGCCAGAAAACACACATCATTTAACTAATCTAAT 5159
|||
5041 ACTGAAGAGACACTCATGATGTTGTTAAGAAACAGATGCTGAGTGTGTGTAACGACA 5100
|||||
5160 ACTGAAGAGACACTCATGATGTTGTTAAGAAACAGATGCTGAGTGTGTGTAACGACA 5219
|||
5101 CTGAATATTTTCTAGGAATTTGCGGAGAGAAATGGGTAGTTAGTATTTCTGGTGACC 5160
|||||
5220 CTGAATATTTTCTAGGAATTTGCGGAGAGAAATGGGTAGTTAGTATTTCTGGTGACC 5279
|||||
5161 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGATG 5220
|||||
5280 CAGTCTATTAAAGAAAGAAATGCTGAATGAGCATGATTTTGAAGTCAAGAGATGATG 5339
|||||
5221 GTCATGAGAAAGAACACCAAGGTCGCAAGCGAGCAAGAAATCCAGAGCAGAAAGATC 5280
|||||
5340 GTCATGAGAAAGAACACCAAGGTCGCAAGCGAGCAAGAAATCCAGAGCAGAAAGATC 5399
|||||
5281 TTCAGGGGCTAGAAATCTGTGTGTATGAGGCTTTCACCAACATGCCCCACAGATCACTG 5340
|||||
5400 TTCAGGGGCTAGAAATCTGTGTGTATGAGGCTTTCACCAACATGCCCCACAGATCACTG 5459
|||||
5341 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5400
|||||
5460 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5519
|||||
5401 GGCACAGGTGTCACCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5460
|||||
5520 GGCACAGGTGTCACCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5579
|||||
5461 TTCATGCAATTTGGGAGATGTTGAGGACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5520
|||||
5580 TTCATGCAATTTGGGAGATGTTGAGGACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5639
|||||
5521 AGTGTAGCACTTACCAAGTGTGAGGACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5580
|||||
5640 AGTGTAGCACTTACCAAGTGTGAGGACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5699
|||||
Qy 5581 AGCCACTAC 5589
|||||
Db 5700 AGCCACTAC 5708
|||||

```

```

RESULT 6
US-09-734-672-1
: Sequence 1, Application US/09734672
: Publication No. US20020183268A1
: GENERAL INFORMATION:
: APPLICANT: Murphy, Patricia D.
: Allen, Antonette C.
: Alvarez, Christopher P.
: Critz, Brenda S.
: Olson, Sheri J.
: Scheller, Denise B.
: Zeng, Bin
: TITLE OF INVENTION: Coding Sequences of the Human
: BRCA1 Gene
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Morgan Lewis & Bockius LLP
: STREET: 1111 Pennsylvania Ave., N.W.
: CITY: Washington
: STATE: District of Columbia
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/734,672
: FILING DATE: 03-Dec-2000

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/966,436  
FILING DATE: 07-NO. US20020183268A1-97  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 12-Feb-96  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5055-02-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: No. US20020183268A1 Relevant  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCAL  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-734-672-1

Query Match 99.98; Score 5581; DB 9; Length 5711;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 5584; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAGTTATCTGCTCTGCTGCTGAGAGTACAAATGCTATATGCTATGACAAA 60  
120 ATGAGTTATCTGCTCTGCTGCTGAGAGTACAAATGCTATATGCTATGACAAA 179  
61 ATCTTACAGTGTCCATCTCTGAGAGTATCAAGAACCTGCTCCAAAGTGTGAC 120  
180 ATCTTACAGTGTCCATCTCTGAGAGTATCAAGAACCTGCTCCAAAGTGTGAC 239  
121 CACATATTTTGCAGATTTTGCATGCTAAACTTCTCAACCAAGAAAGAGGCTTACAG 180  
240 CACATATTTTGCAGATTTTGCATGCTAAACTTCTCAACCAAGAAAGAGGCTTACAG 299  
181 TGTCTTATGTAGAGTATGATATACCAAGAGAGCTTACAGAGTATGAT 240  
300 TGTCTTATGTAGAGTATGATATACCAAGAGAGCTTACAGAGTATGAT 359  
241 CAACCTGTGAGAGCTATGAGAAATCATTTGTCTTTTCAAGCTTGCACAGGTTTGAG 300  
360 CAACCTGTGAGAGCTATGAGAAATCATTTGTCTTTTCAAGCTTGCACAGGTTTGAG 419  
301 TATGCAAGAGCTATATTTTGCAGAAAGAGAAATATCTCTCTGCAACTTCAAAAGAT 360  
420 TATGCAAGAGCTATATTTTGCAGAAAGAGAAATATCTCTCTGCAACTTCAAAAGAT 479  
361 GAAGTTCTATCTCCAAAGTATGAGGCTACAGAAACCGTCCAAAGAGCTTACAGAGT 420  
480 GAAGTTCTATCTCCAAAGTATGAGGCTACAGAAACCGTCCAAAGAGCTTACAGAGT 539  
421 GAACCCGAAAGCTCTCTGAGAGAAACCAAGTCTGAGTCTCAAGCTCTTACCTTGA 480  
540 GAACCCGAAAGCTCTCTGAGAGAAACCAAGTCTGAGTCTCAAGCTCTTACCTTGA 539  
481 ACTGTGAGAACTCTGAGAGCAAGAGCGGATCAACCTTCAAAAGAGCTTGTCTACATT 540  
600 ACTGTGAGAACTCTGAGAGCAAGAGCGGATCAACCTTCAAAAGAGCTTGTCTACATT 659  
541 GAATGGAGATCTGATCTCTGAGAGTACCGTTAATAGGCACTTATGCGATGGGGA 600  
660 GAATGGAGATCTGATCTCTGAGAGTACCGTTAATAGGCACTTATGCGATGGGGA 719  
601 GATCAAGAAATTTTACAAATCACCCCTCAAGAGACGAGATGAATTCATGATTCT 660

720 GATCAAGAAATTTTACAAATCACCCCTCAAGAGACGAGATGAATTCATGATTCT 779  
661 GCAAAAAAGGCTGCTTGTGATTTTCTGAGAGCGGATGTACAAATTAATCAATCA 720  
780 GCAAAAAAGGCTGCTTGTGATTTTCTGAGAGCGGATGTACAAATTAATCAATCA 839  
721 CCCAGTATATGATTTTGAACACCACTGAGAGGCTGAGTGTGAGGATCCGAAAG 780  
840 CCCAGTATATGATTTTGAACACCACTGAGAGGCTGAGTGTGAGGATCCGAAAG 899  
781 TATCAGGATGATCTGTTTCAAACTTGCATGTGAGGATGTGACACAAATTAATCA 840  
900 TATCAGGATGATCTGTTTCAAACTTGCATGTGAGGATGTGACACAAATTAATCA 959  
841 AGCTCATTAAGCATGAGAACAGAGCTTATTAATCTCTCAAAAGCAATTAATGTA 900  
960 AGCTCATTAAGCATGAGAACAGAGCTTATTAATCTCTCAAAAGCAATTAATGTA 1019  
901 AAGGCTGAATCTGTATTAATAAGCAACAGGCTGCTTACGAAGAGGCAACATACAG 960  
1020 AAGGCTGAATCTGTATTAATAAGCAACAGGCTGCTTACGAAGAGGCAACATACAG 1079  
961 TGGGCTGGAAGTAAAGAAACATGATATGATAGCGGATCCGACACAGAAAAAGGTA 1020  
1080 TGGGCTGGAAGTAAAGAAACATGATATGATAGCGGATCCGACACAGAAAAAGGTA 1139  
1021 GATCTGAATGCTATCCCTGCTGTGAGAGAAAGATGAAATTAAGCAAACTGCCATGC 1080  
1140 GATCTGAATGCTATCCCTGCTGTGAGAGAAAGATGAAATTAAGCAAACTGCCATGC 1199  
1081 TCAAGAAATCTGTAGATGATGAGAAATGCTCTGATTAACACTTAATTAAGCAGCATTCAG 1140  
1200 TCAAGAAATCTGTAGATGATGAGAAATGCTCTGATTAACACTTAATTAAGCAGCATTCAG 1259  
1141 AAAGTTAATGATGCTTTTCCAGAAAGTATGAACTGTTAGGTTCTGTAGACTCATGAT 1200  
1260 AAAGTTAATGATGCTTTTCCAGAAAGTATGAACTGTTAGGTTCTGTAGACTCATGAT 1319  
1201 GGGGAGCTCAATCAAAATGCCAAAGTATGATATGATGAGGCTTCAAAATGAGGTAGAT 1260  
1320 GGGGAGCTCAATCAAAATGCCAAAGTATGATATGATGAGGCTTCAAAATGAGGTAGAT 1379  
1261 GAATATCTGCTTCTTACAGAAATATGAGCTTACCTGAGGCTTCAAAATGAGGTAGAT 1320  
1380 GAATATCTGCTTCTTACAGAAATATGAGCTTACCTGAGGCTTCAAAATGAGGTAGAT 1439  
1321 ATATGTAAGAGTAAAGAGTTCACCTCCAAATCAGTATGAGATATTTGAAGACAAATA 1380  
1440 ATATGTAAGAGTAAAGAGTTCACCTCCAAATCAGTATGAGATATTTGAAGACAAATA 1499  
1381 TTTGGGAAAACTATGCGAGAAAGGCAAGGCTCCCAACTTAAGCCATGTAAGTAAAT 1440  
1500 TTTGGGAAAACTATGCGAGAAAGGCAAGGCTCCCAACTTAAGCCATGTAAGTAAAT 1559  
1441 CTATATTAAGAGCATTTTGTACTGAGCCACAGATTAATTAAGAGGCTCCCTCACAAT 1500  
1560 CTATATTAAGAGCATTTTGTACTGAGCCACAGATTAATTAAGAGGCTCCCTCACAAT 1619  
1501 AAATTAAGAGCATTTTGTACTGAGCCACAGATTAATTAAGAGGCTCCCTCACAAT 1560  
1620 AAATTAAGAGCATTTTGTACTGAGCCACAGATTAATTAAGAGGCTCCCTCACAAT 1679  
1561 GCGAGATTTGGCAGTTCAAAAGACTCCTGAATATGATATGAGGAACTAACCAAGGAG 1620  
1680 GCGAGATTTGGCAGTTCAAAAGACTCCTGAATATGATATGAGGAACTAACCAAGGAG 1739  
1621 CAGAATGCTCAAGTGAATATTTACTAATAGGCGATCAGAAATTAATCAAAAGGAT 1680  
1740 CAGAATGCTCAAGTGAATATTTACTAATAGGCGATCAGAAATTAATCAAAAGGAT 1799  
1681 TCTATTCAAGATGAGAAAAATCTTAACCCCAATGAATCACTCAAAAAAGATCTGCTTTC 1740

Db 1800 TCTATTGAGATGAGAAAAATCTTAACCAATAGAACTACTCGAAAAAGATCTGCTTC 1859  
Qy 1741 AAAAGAAAGCTGACCTATATAGCAGCATATAGCAATATGAACTCGAATTAATATC 1800  
Db 1860 AAAACAAAGCTTAACCTATATAGCAGCATATAGCAATATGAACTCGAATTAATATC 1919  
Qy 1801 CACAAATTCAAAAGCACCCTAAAAGAAATAGGCTGAGAGAAAGCTTCTTACCAGCATATT 1860  
Db 1920 CCAATATTCAAAAGCACCCTAAAAGAAATAGGCTGAGAGAAAGCTTCTTACCAGCATATT 1979  
Qy 1861 CATGCCCTTGAAGTACTAGTACTAGTAAATCTAAGCCCACTAATTCTACTGAATTCGA 1920  
Db 1980 CATGCCCTTGAAGTACTAGTACTAGTAAATCTAAGCCCACTAATTCTACTGAATTCGA 2039  
Qy 1921 ATTGATAGTGTCTCTAGCAGTGAAGATTAAGAAAAAGTCAACCAATGCCAGTC 1980  
Db 2040 ATTGATAGTGTCTCTAGCAGTGAAGATTAAGAAAAAGTCAACCAATGCCAGTC 2099  
Qy 1981 AGGCACAGAGAAACCTTACAACTCATGAGAGTAAAGAACCTGCACCTGAGCCAGAG 2040  
Db 2100 AGGCACAGAGAAACCTTACAACTCATGAGAGTAAAGAACCTGCACCTGAGCCAGAG 2159  
Qy 2041 AGTAAACAGCCAAATGACAGAGACAGTAAAGACATGACAGTACTTCTCCAGACTG 2100  
Db 2160 AGTAAACAGCCAAATGACAGAGACAGTAAAGACATGACAGTACTTCTCCAGACTG 2219  
Qy 2101 AAGTTAACAAATGCACCTGCTTCTTACTAAGTGTCAATACCACTGAACCTTAAAGAA 2160  
Db 2220 AAGTTAACAAATGCACCTGCTTCTTACTAAGTGTCAATACCACTGAACCTTAAAGAA 2279  
Qy 2161 TTTGTCAATCTAGCCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2220  
Db 2280 TTTGTCAATCTAGCCTTCCAGAGAGAAAAAGAAAGAACTAGAAACAGTTAAAGTG 2339  
Qy 2221 TCTAAATAGCTGAAACACCCCAAGATCTCATGTATTAAGTGAAGAAAGGTTTGGCAACT 2280  
Db 2340 TCTAAATAGCTGAAACACCCCAAGATCTCATGTATTAAGTGAAGAAAGGTTTGGCAACT 2399  
Qy 2281 GAAAGATCTGAGAGAGTACATATTTCAATGTGTAACCTGTAAGTATGACACTGAG 2340  
Db 2400 GAAAGATCTGAGAGAGTACATATTTCAATGTGTAACCTGTAAGTATGACACTGAG 2459  
Qy 2341 GAAAGATCTGAGAGAGTACATATTTCAATGTGTAACCTGTAAGTATGACACTGAG 2400  
Db 2460 GAAAGATCTGAGAGAGTACATATTTCAATGTGTAACCTGTAAGTATGACACTGAG 2519  
Qy 2401 TGGTAGTGTGAGTGTGAGATTTGAAACCCCAAGGACTAATTCATGTTGTCCAAA 2460  
Db 2520 TGGTAGTGTGAGTGTGAGATTTGAAACCCCAAGGACTAATTCATGTTGTCCAAA 2579  
Qy 2461 GATTAATAGAAATGACACAGAAAGCTTAAAGTATCATTTGGGACATGAAGTTAACACAGT 2520  
Db 2580 GATTAATAGAAATGACACAGAAAGCTTAAAGTATCATTTGGGACATGAAGTTAACACAGT 2639  
Qy 2521 CGGGAAACAGATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCAAGATACA 2580  
Db 2640 CGGGAAACAGATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCAAGATACA 2699  
Qy 2581 TTCAAGGTTTCAAAAGGCGAGCATTTGCTGCTTTTCAATCCAGAAATGGAAGAG 2640  
Db 2700 TTCAAGGTTTCAAAAGGCGAGCATTTGCTGCTTTTCAATCCAGAAATGGAAGAG 2759  
Qy 2641 GAATGTGCAACATTTCTGCCCCACTGCGGCTCTTAAGAAACAAAGTCCAAAAGTCACT 2700  
Db 2760 GAATGTGCAACATTTCTGCCCCACTGCGGCTCTTAAGAAACAAAGTCCAAAAGTCACT 2819  
Qy 2701 TTTGAATGTGAACAAAAGAGAAATCAAGAAAGATGAGTCTAATATCAAGCCCTGA 2760  
Db 2820 TTTGAATGTGAACAAAAGAGAAATCAAGAAAGATGAGTCTAATATCAAGCCCTGA 2879  
Qy 2761 CAGACAGTTAATATCATGCTGAGGCTTCTGCTGTGTTGCTCAGAAAGATTAAGCCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATCATGCTGAGGCTTCTGCTGTGTTGCTCAGAAAGATTAAGCCAGTTGAT 2939

Qy 2821 AATGCCAAATGTAGTATCAAAAGAGGCTCTAGGTTTGTCTATACATCTCAGTTTCAGAGGC 2880  
Db 2940 AATGCCAAATGTAGTATCAAAAGAGGCTCTAGGTTTGTCTATACATCTCAGTTTCAGAGGC 2999  
Qy 2881 AACGAAATGTGACTCATTTACTCTCAAAATTAACATGACCTTTTACAAAAACCATATTCGTATA 2940  
Db 3000 AACGAAATGTGACTCATTTACTCTCAAAATTAACATGACCTTTTACAAAAACCATATTCGTATA 3059  
Qy 2941 CCACACTTTTTCCTCATCAAGTCAATTTGTTAAACCTAATGTAAAGAAAAATCTGCTAGAG 3000  
Db 3060 CCACACTTTTTCCTCATCAAGTCAATTTGTTAAACCTAATGTAAAGAAAAATCTGCTAGAG 3119  
Qy 3001 GAAACCTTTGAGGAACATTCATATGTCACCGTGAAGAAAGTGGAAATGAGAACAATTCGA 3060  
Db 3120 GAAACCTTTGAGGAACATTCATATGTCACCGTGAAGAAAGTGGAAATGAGAACAATTCGA 3179  
Qy 3061 AGTACAGTGAACAATTAAGCCGTAATTAACATTAAGAGAAATGTTTTTAAAGACCCAGC 3120  
Db 3180 AGTACAGTGAACAATTAAGCCGTAATTAACATTAAGAGAAATGTTTTTAAAGACCCAGC 3239  
Qy 3121 TCAAGCAATTTATATGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAAATGA 3180  
Db 3240 TCAAGCAATTTATATGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAAATGA 3299  
Qy 3181 ATAGGTTCCAGTATGAAAAACATTCAGAGACAATAGTGAAGACAAGAGGCCAAATTTG 3240  
Db 3300 ATAGGTTCCAGTATGAAAAACATTCAGAGACAATAGTGAAGACAAGAGGCCAAATTTG 3359  
Qy 3241 AATGCTATGCTTATGATTAAGGGGTTTTCGAACTGAGCTGTATTAACAAAGTCTTCTGGA 3300  
Db 3360 AATGCTATGCTTATGATTAAGGGGTTTTCGAACTGAGCTGTATTAACAAAGTCTTCTGGA 3419  
Qy 3301 AGTAAATTTAGATCTCTGAAATTAAGAAAGCAATTAAGAAAGTACTGACTGCT 3360  
Db 3420 AGTAAATTTAGATCTCTGAAATTAAGAAAGCAATTAAGAAAGTACTGACTGCT 3479  
Qy 3361 AATACAGATTTCTCTCATATCTGATTCAGATTAATCTTGAACACAGCTTATGGAAGTAGT 3420  
Db 3480 AATACAGATTTCTCTCATATCTGATTCAGATTAATCTTGAACACAGCTTATGGAAGTAGT 3539  
Qy 3421 CATGATCTGAGGTTTGTCTGAGACACTGATGACCTGTATGATGATGATGTAATTAAG 3480  
Db 3540 CATGATCTGAGGTTTGTCTGAGACACTGATGACCTGTATGATGATGATGTAATTAAG 3599  
Qy 3481 GAAGATTAAGTGTGCTGAAAAATGACATTAAGGAAAGTGTGCTTTTGAACAAAGC 3540  
Db 3600 GAAGATTAAGTGTGCTGAAAAATGACATTAAGGAAAGTGTGCTTTTGAACAAAGC 3659  
Qy 3541 GTCCAGAAAGGAGAGCTTAGCAGGAGTCTAGCCCTTTCACCCATTAACATTTGGCTCAG 3600  
Db 3660 GTCCAGAAAGGAGAGCTTAGCAGGAGTCTAGCCCTTTCACCCATTAACATTTGGCTCAG 3719  
Qy 3601 GGTACCAGAAAGGAGGCGCAAGAAATTAAGTCTCAGAAAGAACTATCTAGTGAAGAT 3660  
Db 3720 GGTACCAGAAAGGAGGCGCAAGAAATTAAGTCTCAGAAAGAACTATCTAGTGAAGAT 3779  
Qy 3661 GAAGAGCTTCCGCTTCCACACCTGTTATTTGTAAGTAAACAATATCCTTCTCAG 3720  
Db 3780 GAAGAGCTTCCGCTTCCACACCTGTTATTTGTAAGTAAACAATATCCTTCTCAG 3839  
Qy 3721 TCTACTAGGCAATAGCACCGCTTGTACCGAGTCTGTCTAAGAAACACAGAGAGAAATTTA 3780  
Db 3840 TCTACTAGGCAATAGCACCGCTTGTACCGAGTCTGTCTAAGAAACACAGAGAGAAATTTA 3899  
Qy 3781 TTATCAATTAAGAAATAGCTTAAATGACATGCAATTAACAGATTAATTTGGCAAGGCAATCT 3840  
Db 3900 TTATCAATTAAGAAATAGCTTAAATGACATGCAATTAACAGATTAATTTGGCAAGGCAATCT 3959  
Qy 3841 CAGGAACATCACCTTGTGAGAGAAACAAATGTCTGAGCTGTGTTTCTTCCACAGTGC 3900  
Db 3960 CAGGAACATCACCTTGTGAGAGAAACAAATGTCTGAGCTGTGTTTCTTCCACAGTGC 4019

QY 3901 AGTGAATGGAGACTGCTGCAATACAAACCCAGAGATCCTTTCTGATTTGTTCT 3960  
 Db 4020 AGTGAATGGAGACTGCTGCAATACAAACCCAGAGATCCTTTCTGATTTGTTCT 4079  
 QY 3961 TCCAAACAAATGAGGATCAGTCTGAAAGCCAGGAGTTGTTGATGACAAAGAAATTG 4020  
 Db 4080 TCCAAACAAATGAGGATCAGTCTGAAAGCCAGGAGTTGTTGATGACAAAGAAATTG 4139  
 QY 4021 GTTTCAGATGATGAAG 4080  
 Db 4140 GTTTCAGATGATGAAG 4199  
 QY 4081 ATGATTTCAAACTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
 Db 4200 ATGATTTCAAACTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4259  
 QY 4141 GATGCTCAGAGGCTATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
 Db 4260 GATGCTCAGAGGCTATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4319  
 QY 4201 CAACATTAACCTGATTAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260  
 Db 4320 CAACATTAACCTGATTAAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4379  
 QY 4261 CATGGAG 4320  
 Db 4380 CATGGAG 4439  
 QY 4321 GACTCGGAATCCAGAAACAG 4380  
 Db 4440 GACTCGGAATCCAGAAACAG 4499  
 QY 4381 AGTGAATACCTTAAGCTTCAG 4440  
 Db 4500 AGTGAATACCTTAAGCTTCAG 4559  
 QY 4441 GCAGATATGTTCTACAGTAAATTAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
 Db 4560 GCAGATATGTTCTACAGTAAATTAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4619  
 QY 4501 TGCCCATCATTAAGTGAATGAGTGTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
 Db 4620 TGCCCATCATTAAGTGAATGAGTGTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4679  
 QY 4561 AACTACCATCTCAG 4620  
 Db 4680 AACTACCATCTCAG 4739  
 QY 4621 GAGTCTGGGCGACAGATTTGACGGAAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA 4680  
 Db 4740 GAGTCTGGGCGACAGATTTGACGGAAACATCTTACTTGGCAAGGCAAGATCTAGAGGGA 4799  
 QY 4681 ACCCTTACCTGGAATGGAATCAGCTCTTCTCTATGACCTGGAATCTGATCTTCT 4740  
 Db 4800 ACCCTTACCTGGAATGGAATCAGCTCTTCTCTATGACCTGGAATCTGATCTTCT 4859  
 QY 4741 GAAGACAG 4800  
 Db 4860 GAAGACAG 4919  
 QY 4801 AAAGTTCCCAATTTGAAAGTTGCAAGATCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
 Db 4920 AAAGTTCCCAATTTGAAAGTTGCAAGATCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4979  
 QY 4861 GATACCTGCTGGATATATGCAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
 Db 4980 GATACCTGCTGGATATATGCAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5039  
 QY 4921 GCTTCAACAGAAAGGCTCAACAAAG 4980  
 Db 5040 GCTTCAACAGAAAGGCTCAACAAAG 5099  
 QY 4981 GAATTTATGCTGCTGTAAGAGTTGGCAAGAAACACCAATCACTTAATTAAT 5040

Db 5100 GAATTTATGCTGCTGTAAGAGTTGGCAAGAAACACCAATCACTTAATTAATTAAT 5159  
 QY 5041 ACTGAAG 5100  
 Db 5160 ACTGAAG 5219  
 QY 5101 CTGAATATTTTCTAG 5160  
 Db 5220 CTGAATATTTTCTAG 5279  
 QY 5161 CAGTCTATTTAAG 5220  
 Db 5280 CAGTCTATTTAAG 5339  
 QY 5221 GTCAATGGAAGAAACCCAG 5280  
 Db 5340 GTCAATGGAAGAAACCCAG 5399  
 QY 5281 TTCAAGGAG 5340  
 Db 5400 TTCAAGGAG 5459  
 QY 5341 GAATGATGATGAG 5400  
 Db 5460 GAATGATGATGAG 5519  
 QY 5401 GGCACAG 5460  
 Db 5520 GGCACAG 5579  
 QY 5461 TTCCATGCAATTTGGGAG 5520  
 Db 5580 TTCCATGCAATTTGGGAG 5639  
 QY 5521 AGTGTACACTCTACAG 5580  
 Db 5640 AGTGTACACTCTACAG 5699  
 QY 5581 AGCCACTAC 5589  
 Db 5700 AGCCACTAC 5708

RESULT 7  
 US-09-982-828-1  
 : Sequence 1, Application US/09982828  
 : Publication No. US20030022184A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Murphy, Patricia D.  
 : Allen, Antonette C.  
 : Alvarez, Christopher P.  
 : Critz, Brenda S.  
 : Olson, Sheri J.  
 : Thurber, Denise  
 : Zeng, Bin  
 : TITLE OF INVENTION: Coding Sequences of the Human  
 : BRCAL Gene  
 : NUMBER OF SEQUENCES: 72  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Morgan Lewis & Bockius LLP  
 : STREET: 1111 Pennsylvania Avenue N. W.  
 : CITY: Washington  
 : STATE: DC  
 : COUNTRY: USA  
 : ZIP: 20004  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/982,828



FILED DATE: 22-Oct-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/074,453  
FILING DATE: 1998-05-06  
APPLICATION NUMBER: US 08/798,691  
FILING DATE: 1997-02-12  
APPLICATION NUMBER: US 08/598,591  
FILING DATE: 1996-02-12  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael S. Tuscan  
REGISTRATION NUMBER: 43,210  
REFERENCE/DOCKET NUMBER: 44921-5053-01-05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-739-3000  
TELEFAX: 202-739-3001  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
STRAIN: BRCA1 (om11)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 17  
MAP POSITION: 17q21  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-982-828-1

Query Match 99.9%; Score 5581; DB 9; Length 5711;  
Best local Similarity 99.9%; Pred. No. 0;  
Matches 5584; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGGATTTATCTGCTCTTGGCTTGAGAAAGTCAATTAATGCTATGACAAA 60  
DB 120 ATGGATTTATCTGCTCTTGGCTTGAGAAAGTCAATTAATGCTATGACAAA 179  
QY 61 ATCTTGAGTGTCCATCTGCTGAGAGTGTGATCAAGAACCTGCTCCACAAGTGTGAC 120  
DB 180 ATCTTGAGTGTCCATCTGCTGAGAGTGTGATCAAGAACCTGCTCCACAAGTGTGAC 239  
QY 121 CACATATTTTGGCAATTTTGCATGTGAACCTCTCAACGAGAAGAGGCGCTTCACAG 180  
DB 240 CACATATTTTGGCAATTTTGCATGTGAACCTCTCAACGAGAAGAGGCGCTTCACAG 299  
QY 181 TGTCTTTATGTAGATGATATACCAAAAGAGCCTACAGAAAGTACGAGATTTAGT 240  
DB 300 TGTCTTTATGTAGATGATATATACCAAAAGAGCCTACAGAAAGTACGAGATTTAGT 359  
QY 241 CAACCTGTGAAGAGCTATGAAATCATTTGCTTTTCAGCTTGACAGGTTGGAG 300  
DB 360 CAACCTGTGAAGAGCTATGAAATCATTTGCTTTTCAGCTTGACAGGTTGGAG 419  
QY 301 TATGCAAAAGCATAATTTTGGCAAAAAGAAATTAACCTCTCTGAACATCTAAAAGAT 360  
DB 420 TATGCAAAAGCATAATTTTGGCAAAAAGAAATTAACCTCTCTGAACATCTAAAAGAT 479  
QY 361 GAAGTTTCTATCATCAAAAGATAGGCTACAGAAACCGTCCAAAAGACTTCTACAGAGT 420  
DB 480 GAAGTTTCTATCATCAAAAGATAGGCTACAGAAACCGTCCAAAAGACTTCTACAGAGT 539  
QY 421 GAACCGGAAATTCCTCTTGGCAAAACAGCTCAGTGTCCAACTCTTAACCTTGA 480  
DB 540 GAACCGGAAATTCCTCTTGGCAAAACAGCTCAGTGTCCAACTCTTAACCTTGA 599  
QY 481 ACTGTGAGAACTCTGAGCAAAAGAGCGATACAACTCAAAAGAGCTGTGCTACATT 540  
DB 600 ACTGTGAGAACTCTGAGCAAAAGAGCGATACAACTCAAAAGAGCTGTGCTACATT 659  
QY 541 GAATTGGATCTGATTTCTTCTGAAGATACCGTTAATTAAGCAACTTAATTCAGTGTGGA 600

DB 660 GAATTGGATCTGATTTCTTCTGAAGATACCGTTAATTAAGCAACTTAATTCAGTGTGGA 719  
QY 601 GATCAACAATTTGTACAAATCACCCTCAAGAGAACGAGGATGAATCGTTGGATCT 660  
DB 720 GATCAACAATTTGTACAAATCACCCTCAAGAGAACGAGGATGAATCGTTGGATCT 779  
QY 661 GCAAAAAGAGCTCTTGTGAATTTTCTGAGAGCGATGTAAACAATACTGACATCA 720  
DB 780 GCAAAAAGAGCTCTTGTGAATTTTCTGAGAGCGATGTAAACAATACTGACATCA 839  
QY 721 CCCAGTATATATGATTTGAACACCACTGAGAACGCTGACGCTGAGAGCATCCAGAAAG 780  
DB 840 CCCAGTATATATGATTTGAACACCACTGAGAACGCTGACGCTGAGAGCATCCAGAAAG 899  
QY 781 TATCAGGCTGCTCTGTTTCAAACTGCATGTGGAGCCATGTGGCAACAATCTCATGCC 840  
DB 900 TATCAGGCTGCTCTGTTTCAAACTGCATGTGGAGCCATGTGGCAACAATCTCATGCC 959  
QY 841 AGCTCATTTACAGCATGAGAACAGCACTTATTACTACTTAAGACAGAAATGATAGAA 900  
DB 960 AGCTCATTTACAGCATGAGAACAGCACTTATTACTACTTAAGACAGAAATGATAGAA 1019  
QY 901 AAGCTGATTTCTGTAAATTAAGCAAAACAGCTGCTTACGAAAGAGCCACATTAACGA 960  
DB 1020 AAGCTGATTTCTGTAAATTAAGCAAAACAGCTGCTTACGAAAGAGCCACATTAACGA 1079  
QY 961 TGGGCTGGAATGAGGAAACATGTATGATGAGCGGACCTCCAGCAGCAAAAAAAGGA 1020  
DB 1080 TGGGCTGGAATGAGGAAACATGTATGATGAGCGGACCTCCAGCAGCAAAAAAAGGA 1139  
QY 1021 GATCTGATGATGATCCCTGTGTGAGAGAAAGAAATGATGATGAGCAAACTGCATGC 1080  
DB 1140 GATCTGATGATGATCCCTGTGTGAGAGAAAGAAATGATGATGAGCAAACTGCATGC 1199  
QY 1081 TCAGAGATCTTGAGATGATGAGATGTTCTTGTGATTAACATAATAGCAGATTGAG 1140  
DB 1200 TCAGAGATCTTGAGATGATGAGATGTTCTTGTGATTAACATAATAGCAGATTGAG 1259  
QY 1141 AAAGTTAATGATGATGTTTTCAGAGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 1260 AAAGTTAATGATGATGTTTTCAGAGATGATGATGATGATGATGATGATGATGATGAT 1319  
QY 1201 GGGGAGCTGGAATCAATATGCCAAGTAGCTGATGATGATGATGATGATGATGATGAT 1260  
DB 1320 GGGGAGCTGGAATCAATATGCCAAGTAGCTGATGATGATGATGATGATGATGATGATGAT 1379  
QY 1261 GAATATTTCTGTTCTTCAAGAAATAGACTTACTGCGCAGATGATCTCATGAGCTTTA 1320  
DB 1380 GAATATTTCTGTTCTTCAAGAAATAGACTTACTGCGCAGATGATCTCATGAGCTTTA 1439  
QY 1321 ATATGTTAAAGTGAAGAGTTCACCTCAATCTGATGAGATGATGATGATGATGATGAT 1380  
DB 1440 ATATGTTAAAGTGAAGAGTTCACCTCAATCTGATGAGATGATGATGATGATGATGAT 1499  
QY 1381 TTTGGGAAAACCTATCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1500 TTTGGGAAAACCTATCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559  
QY 1441 CTAATTTATAGAGCATTTTGTACTGAGCCAGATATATCAAGAGCGTCCCTCACAAT 1500  
DB 1560 CTAATTTATAGAGCATTTTGTACTGAGCCAGATATATCAAGAGCGTCCCTCACAAT 1619  
QY 1501 AAATTTAAAGCTAAAG 1560  
DB 1620 AAATTTAAAGCTAAAG 1679  
QY 1561 GCAGATTTGGCAATTCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
DB 1680 GCAGATTTGGCAATTCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1739  
QY 1621 CAGATGCTCAAGTGAATATTAATTAATGATGATGATGATGATGATGATGATGATGAT 1680



Db 1740 CAAATGCTCAATGATGATATTACTAATAGTGTCATGAGATTAACAAAAGGTGAT 1799  
QY 1681 TCTATTGAGATAGAGAAAATCCTTAACCAATAGAAATCACTCGAAAAAGAAATGCTTTC 1740  
Db 1800 TCTATTGAGATAGAGAAAATCCTTAACCAATAGAAATCACTCGAAAAAGAAATGCTTTC 1859  
QY 1741 AAAACGAAAGCTGACCTATTAAGCAGCAGTAAAGCAATATGGAAGTGAATTAATATC 1800  
Db 1860 AAAACGAAAGCTGACCTATTAAGCAGCAGTAAAGCAATATGGAAGTGAATTAATATC 1919  
QY 1801 CACAATTCAAAGACACTTAAAGAAATGAGGAGGAGAACTTCTACAGCAGATAT 1860  
Db 1920 CACAATTCAAAGACACTTAAAGAAATGAGGAGGAGAACTTCTACAGCAGATAT 1979  
QY 1861 CATGCGCTTGAAGTACTAGTACAGTAAAGTAAAGCCACTAATGTACTGATTTGCA 1920  
Db 1980 CATGCGCTTGAAGTACTAGTACAGTAAAGTAAAGCCACTAATGTACTGATTTGCA 2039  
QY 1921 ATTGATATGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCCAGTC 1980  
Db 2040 ATTGATATGTTGTTCTAGCAGTGAAGATTAAGAAAAAAGTACAAACCAATGCCAGTC 2099  
QY 1981 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGAACTCGCACTGAGCCAGAG 2040  
Db 2100 AGGCACAGCAGAAACCTACACACTCATGGAAGTAAAGAACTCGCACTGAGCCAGAG 2159  
QY 2041 AGTACACAGCAGCAATGACAGCAGTAAAGACATGACAGTATCTTCCAGAGCTG 2100  
Db 2160 AGTACACAGCAGCAATGACAGCAGTAAAGACATGACAGTATCTTCCAGAGCTG 2219  
QY 2101 AAGTTAAACAATGACACCTGCTCTTTTACTAAGTGTCAAAATACAGTGAATTAAGAA 2160  
Db 2220 AAGTTAAACAATGACACCTGCTCTTTTACTAAGTGTCAAAATACAGTGAATTAAGAA 2279  
QY 2161 TTTGTAATCCCTTGAAGCTTCCAGAGAGAAAAAGAGAGAACTACAGAACTTAAGT 2220  
Db 2280 TTTGTAATCCCTTGAAGCTTCCAGAGAGAAAAAGAGAGAACTACAGAACTTAAGT 2339  
QY 2221 TCTAATATGCTGAAAGACCCCAAGATCTCATGTAAAGTGAAGAGGTTTGCAGACT 2280  
Db 2340 TCTAATATGCTGAAAGACCCCAAGATCTCATGTAAAGTGAAGAGGTTTGCAGACT 2399  
QY 2281 GAAAGATCTGTAGAGAGTACAGTATTTCAATGCTGTAAGTATGACACTCAG 2340  
Db 2400 GAAAGATCTGTAGAGAGTACAGTATTTCAATGCTGTAAGTATGACACTCAG 2459  
QY 2341 GAAAGATCTGTAGAGAGTACAGTATTTCAATGCTGTAAGTATGACACTCAG 2400  
Db 2460 GAAAGATCTGTAGAGAGTACAGTATTTCAATGCTGTAAGTATGACACTCAG 2519  
QY 2401 TGTGTAGTCACTGTGACAGATTTGAAAACCCCAAGGAGCTAATTCATGTTGTTCCAA 2460  
Db 2520 TGTGTAGTCACTGTGACAGATTTGAAAACCCCAAGGAGCTAATTCATGTTGTTCCAA 2579  
QY 2461 GATATAGAAATGACACAGAGGCTTTAAGTATCCATGAGACATGAAGTTAACACAGT 2520  
Db 2580 GATATAGAAATGACACAGAGGCTTTAAGTATCCATGAGACATGAAGTTAACACAGT 2639  
QY 2521 CGGGAACAAGCATTAAGATGGAAGAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2580  
Db 2640 CGGGAACAAGCATTAAGATGGAAGAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2699  
QY 2581 TTCAGAGTTCAAGAGGCGAGTATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2640  
Db 2700 TTCAGAGTTCAAGAGGCGAGTATTTGCTGTTTCAATTCAGAGAAATGAGAGAG 2759  
QY 2641 GAATGTGCAACATTTCTGCTCCCACTGTGGTCTTTAAGAAACCAAGTCCAAAGTCACT 2700  
Db 2760 GAATGTGCAACATTTCTGCTCCCACTGTGGTCTTTAAGAAACCAAGTCCAAAGTCACT 2819  
QY 2701 TTTGATGTGAACAAAGAGAGAAATTCAGAGAAATGAGTCTATATCAAGCTGTGA 2760  
Db 2820 TTTGATGTGAACAAAGAGAGAAATTCAGAGAAATGAGTCTATATCAAGCTGTGA 2879

QY 2761 CAGACAGTTAATATACATGAGGCTTTCCTGTGTTGTCAGAAAGATTAAGCCAGTTGAT 2820  
Db 2880 CAGACAGTTAATATACATGAGGCTTTCCTGTGTTGTCAGAAAGATTAAGCCAGTTGAT 2939  
QY 2821 AATGCCAATATGATCAAGAGAGGCTTTGCTGTATCATCTCAGTTCAAGAGG 2880  
Db 2940 AATGCCAATATGATCAAGAGAGGCTTTGCTGTATCATCTCAGTTCAAGAGG 2999  
QY 2881 AAGGAACCTGACTCATTTACTCCAAATTAACATGAGCTTTTACAAAACCATATCGATA 2940  
Db 3000 AAGGAACCTGACTCATTTACTCCAAATTAACATGAGCTTTTACAAAACCATATCGATA 3059  
QY 2941 CCACACCTTTTCCCATCAAGTCAATTTGTTTAAAGTAAATGTAAAGAAATTCGTAGAG 3000  
Db 3060 CCACACCTTTTCCCATCAAGTCAATTTGTTTAAAGTAAATGTAAAGAAATTCGTAGAG 3119  
QY 3001 GAAACCTTTGAGAGACATTTCAATGTCACTGAAAGAGAAATGGAATGGAACATTTCA 3060  
Db 3120 GAAACCTTTGAGAGACATTTCAATGTCACTGAAAGAGAAATGGAATGGAACATTTCA 3179  
QY 3061 AGTACAGTGAACCAATTTAGCCGTATTAACATTAGAGAAATGTTTAAAGAACCCAGC 3120  
Db 3180 AGTACAGTGAACCAATTTAGCCGTATTAACATTAGAGAAATGTTTAAAGAACCCAGC 3239  
QY 3121 TCAAGCAATTTATGAGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3180  
Db 3240 TCAAGCAATTTATGAGAGTGTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGA 3299  
QY 3181 ATAGGTTCCAGTATGAGAGAAACATTTCAAGAGAACTAGTAAAGAGAGGCCCAAAATG 3240  
Db 3300 ATAGGTTCCAGTATGAGAGAAACATTTCAAGAGAACTAGTAAAGAGAGGCCCAAAATG 3359  
QY 3241 AATGCTATGCTTATGATTTAGGAGGTTTGCACACCTGAGGCTATTAACAAAGTCTTCTGCA 3300  
Db 3360 AATGCTATGCTTATGATTTAGGAGGTTTGCACACCTGAGGCTATTAACAAAGTCTTCTGCA 3419  
QY 3301 AGTAAATTTGAGAGATCCCTGAAATTAAGAAACAGAAATATGAAGAGTTCAGACTGT 3360  
Db 3420 AGTAAATTTGAGAGATCCCTGAAATTAAGAAACAGAAATATGAAGAGTTCAGACTGT 3479  
QY 3361 AATACAGATTTCTCTCCATATCTGATTTAGATTAAGACAGCCTATGGAAGTATG 3420  
Db 3480 AATACAGATTTCTCTCCATATCTGATTTAGATTAAGACAGCCTATGGAAGTATG 3539  
QY 3421 CATGCAATCTCAGGTTTCTGTGACACCTGATGACCTGTTAGATGATGATGAATTAAG 3480  
Db 3540 CATGCAATCTCAGGTTTCTGTGACACCTGATGACCTGTTAGATGATGATGAATTAAG 3599  
QY 3481 GAAGATAGTACTGTTTCTGAGAAATGACATTAAGAGAAATGCTGCTTTTACCAAAAGC 3540  
Db 3600 GAAGATAGTACTGTTTCTGAGAAATGACATTAAGAGAAATGCTGCTTTTACCAAAAGC 3659  
QY 3541 GTCCAGAAAGAGAGGCTTAGCAGAGGCTTACAGCCCTTTACACCATTTGAGCTCAG 3600  
Db 3660 GTCCAGAGAGAGGCTTAGCAGAGGCTTACAGCCCTTTACACCATTTGAGCTCAG 3719  
QY 3601 GGTACCGAGAGAGGCGCAAGAAATTAAGTCTCTGAGAGAGAACTTATCTAGTGAAT 3660  
Db 3720 GGTACCGAGAGAGGCGCAAGAAATTAAGTCTCTGAGAGAGAACTTATCTAGTGAAT 3779  
QY 3661 GAAGAGCTTCCCGGCTTCCACACCTGTTATTTGTAAGTAAACAAATTTACTCTCTCAG 3720  
Db 3780 GAAGAGCTTCCCGGCTTCCACACCTGTTATTTGTAAGTAAACAAATTTACTCTCTCAG 3839  
QY 3721 TCTACTAGGATATGACACCGTGTCTACAGAGTCTGTCTTAAGAACACAGAGAGAAATTA 3780  
Db 3840 TCTACTAGGATATGACACCGTGTCTACAGAGTCTGTCTTAAGAACACAGAGAGAAATTA 3899  
QY 3781 TTATCATTTAAGAAATATGCTTAATATGATGCTGATTAACAGGTTATTTGCAAAAGCATCT 3840  
Db 3900 TTATCATTTAAGAAATATGCTTAATATGATGCTGATTAACAGGTTATTTGCAAAAGCATCT 3959

3841 CAGGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCACAGTGC 3900  
3960 CAGGAACATCACCTTAGTGAAGAAACAAATGTTCTGCTAGCTGTTTCTTCACAGTGC 4019  
3901 AGTGAATTTGGAACACTTGAATGCAATACAAACACCCAGATGCTTCTTGATTGTTCT 3960  
4020 AGTGAATTTGGAACACTTGAATGCAATACAAACACCCAGATGCTTCTTGATTGTTCT 4079  
3961 TCCAAACAAATGAGGATCAGTCTGAAAGCCAGGAGTGTGCTGATGACAAAGAAATG 4020  
4080 TCCAAACAAATGAGGATCAGTCTGAAAGCCAGGAGTGTGCTGATGACAAAGAAATG 4139  
4021 GTTTCAGATGATGAAGAAAGAGAGGCGCTTGGAAGAAATATCAAGAGAGCAAGC 4080  
4140 GTTTCAGATGATGAAGAAAGAGAGGCGCTTGGAAGAAATATCAAGAGAGCAAGC 4199  
4081 ATGAGATTCAAACTAGTGAAGAGATCTGGGTGTGAAGTGAACCAACCGTCTCGAA 4140  
4200 ATGAGATTCAAACTAGTGAAGAGATCTGGGTGTGAAGTGAACCAACCGTCTCGAA 4259  
4141 GACTGCTCAGGGCTATCTCTCTCAGAGTGAATTTAACTCAGCAGAGGATACCATG 4200  
4260 GACTGCTCAGGGCTATCTCTCTCAGAGTGAATTTAACTCAGCAGAGGATACCATG 4319  
4201 CAACATACCTGATTAAGCTCCAGCAGAGAAATGGCTGAAGCTGTGTAGAACAG 4260  
4320 CAACATACCTGATTAAGCTCCAGCAGAGAAATGGCTGAAGCTGTGTAGAACAG 4379  
4261 CATGGAGGCGAGCTTCTAACAGTACCTTCATCATAGTGAATCTTGGCCCTTGAG 4320  
4380 CATGGAGGCGAGCTTCTAACAGTACCTTCATCATAGTGAATCTTGGCCCTTGAG 4439  
4321 GACTGCGAATTCAGAAACAAACACATCAGAAAGCAATTAATCTTACAGAAAGT 4380  
4440 GACTGCGAATTCAGAAACAAACACATCAGAAAGCAATTAATCTTACAGAAAGT 4499  
4381 AGTGAATACCTTAATGAACAGAAATCCAGAAAGGCTTCTGCTGACAAATTTGAGGTCT 4440  
4500 AGTGAATACCTTAATGAACAGAAATCCAGAAAGGCTTCTGCTGACAAATTTGAGGTCT 4559  
4441 GCAGATAGTTCTACAGTAAATTAAGAACAGAGAGTGAAGGTCATCCCTTCATAA 4500  
4560 GCAGATAGTTCTACAGTAAATTAAGAACAGAGAGTGAAGGTCATCCCTTCATAA 4619  
4501 TGCCCATCATTTAGATGATAGTGTGATGACAGTGTGCTGCGAGATCTTCAGAAATGA 4560  
4620 TGCCCATCATTTAGATGATAGTGTGATGACAGTGTGCTGCGAGATCTTCAGAAATGA 4679  
4561 AACTACCCATCTCAGAGAGAGCTCATTAAAGTTGTTGATGTGAGAGACACAGCTGGAA 4620  
4680 AACTACCCATCTCAGAGAGAGCTCATTAAAGTTGTTGATGTGAGAGACACAGCTGGAA 4739  
4621 GAGTCTGGGCAACAGATTTGACGGAACATCTTACCTGGCAGGGAAGTCTAGAGGGA 4680  
4740 GAGTCTGGGCAACAGATTTGACGGAACATCTTACCTGGCAGGGAAGTCTAGAGGGA 4799  
4681 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGAGACCCCTAATCTGATCTTCT 4740  
4800 ACCCTTACCTGGAATCTGGAATCAGCCCTTCTGATGAGACCCCTAATCTGATCTTCT 4859  
4741 GAAAGACAGAGCCAGAGTCACTGTTGGCAGACATACCTTCAACCTCTGCATTTG 4800  
4860 GAAAGACAGAGCCAGAGTCACTGTTGGCAGACATACCTTCAACCTCTGCATTTG 4919  
4801 AAGGTTCCTCAATTTGAAGTTCAGAAATCTGCCCAAGTCCAGCTCTCTCTACTACT 4860  
4920 AAGGTTCCTCAATTTGAAGTTCAGAAATCTGCCCAAGTCCAGCTCTCTCTACTACT 4979  
4861 GATACCTGCTGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGGAGCAGAAATGACA 4920  
4980 GATACCTGCTGGTATATGCAATGGAAGAAAGTGTGAGCAGGAGGAGCAGAAATGACA 5039  
4921 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGTGCTGACCCAGAA 4980

5040 GCTTCAACAGAAAGGCTCAACAAAGAAATGTCATGCTGTGTCGCTGACCCAGAA 5099  
4981 GAATTTATGCTGCTGTACAAAGTTTGGCAGAAACACCATATCACTTAATATTAAT 5040  
5100 GAATTTATGCTGCTGTACAAAGTTTGGCAGAAACACCATATCACTTAATATTAAT 5159  
5041 ACTGAAGAGACTACTATGTTGTATGAAGACAGATGCTGAGCTTGTGTGACAGGACA 5100  
5160 ACTGAAGAGACTACTATGTTGTATGAAGACAGATGCTGAGCTTGTGTGACAGGACA 5219  
5101 CTGAATATTTCTTACGAATTCGCGAGGAAATGGGTAGTATTTCTGCGGTGACC 5160  
5220 CTGAATATTTCTTACGAATTCGCGAGGAAATGGGTAGTATTTCTGCGGTGACC 5279  
5161 CAGTCTATTAAAGAAAGAAATGCTGATGACCATGATTTTGAGAGTCAAGAGAAATG 5220  
5280 CAGTCTATTAAAGAAAGAAATGCTGATGACCATGATTTTGAGAGTCAAGAGAAATG 5339  
5221 GTCAATGAGAAAGAACCAAGGTCCAAGCAGCAGAGAAATCCAGAGCAGAAAGATC 5280  
5340 GTCAATGAGAAAGAACCAAGGTCCAAGCAGCAGAGAAATCCAGAGCAGAAAGATC 5399  
5281 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAATGCTCCAGATCACTG 5340  
5400 TTCAGGGGCTAGAAATCTGTTGCTATGAGGCTTTCACCAATGCTCCAGATCACTG 5459  
5341 GAATGATGATACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5400  
5460 GAATGATGATACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5519  
5401 GGCACAGGTGTCACCAATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5460  
5520 GGCACAGGTGTCACCAATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5579  
5461 TTCATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGGTGTTGAC 5520  
5580 TTCATGCAATTTGGCAGATGTGTGAGGACCTGTGTGACCCGAGAGTGGGTGTTGAC 5639  
5521 AGTGTACACTGCTACAGTGCAGAGGCTGAGACCTGCTGATACCCACATCCCTCAC 5580  
5640 AGTGTACACTGCTACAGTGCAGAGGCTGAGACCTGCTGATACCCACATCCCTCAC 5699  
5581 AGCCACTAC 5589  
5700 AGCCACTAC 5708

RESULT 8  
US-09-911-904-127

; Sequence 127, Application US/09911904  
; Publication No. US20030096234A1  
; GENERAL INFORMATION:  
; APPLICANT: Farr, Spencer B.  
; APPLICANT: Pickett, Gavin G.  
; APPLICANT: Neft, Robin Eileen  
; APPLICANT: Dunn, II, Robert Thomas  
; TITLE OF INVENTION: CANINE TOXICITY GENES  
; FILE REFERENCE: 4007/42000200  
; CURRENT APPLICATION NUMBER: US/09/911,904  
; CURRENT FILING DATE: 2002-04-09  
; PRIOR APPLICATION NUMBER: US 60/220,057  
; PRIOR FILING DATE: 2000-07-21  
; NUMBER OF SEQ. ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ. ID NO 127  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Canis familiaris  
US-09-911-904-127

Query Match 6.5%; Score 364.6; DB 9; Length 499;  
Best Local Similarity 86.7%; Pred. No. 4,4e-81;



PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 21299  
LENGTH: 147  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO L78833.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3  
OTHER INFORMATION: EST HUMAN HIT: AUI25312.1, EVALUE 2.00e-77  
OTHER INFORMATION: SWISSPROT HIT: P38398, EVALUE 4.00e-12  
OTHER INFORMATION: NT HIT: g111424969, EVALUE 1.00e-77  
US-09-864-761-21299

Query Match 2.6%; Score 147; DB 10; Length 147;  
Best Local Similarity 100.0%; Pred. No. 7.6e-27;  
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4211 TGATAAAGCTCCAGCAGGAATGCTGAGTGAAGCTGTGTTAGAAGCATGGAGCC 4270  
1 TGATAAAGCTCCAGCAGGAATGCTGAGTGAAGCTGTGTTAGAAGCATGGAGCC 60  
Qy 4271 AGCCTTTACAGCAGTACCCCTTCATCATATAGTACTCTTCTGCCCTTGAGACCTGCGAA 4330  
|||||  
Db 61 AGCCTTTACAGCAGTACCCCTTCATCATATAGTACTCTTCTGCCCTTGAGACCTGCGAA 120  
4331 ATCCGAGCAAGACATCAGAAAAG 4357  
|||||  
Qy 121 ATCCGAGCAAGACATCAGAAAAG 147

RESULT 11  
US-09-818-875-661  
Sequence 661, Application US/09818875  
Publication No. US20030051270A1  
GENERAL INFORMATION:  
APPLICANT: Kmiec, Eric B.  
APPLICANT: Gamper, Howard B.  
APPLICANT: Rice, Michael C.  
TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
TITLE OF INVENTION: Stranded Oligonucleotides  
FILE REFERENCE: Napro-4  
CURRENT APPLICATION NUMBER: US/09/818,875  
CURRENT FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: US 60/192,176  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/192,179  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/208,538  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: US 60/244,989  
PRIOR FILING DATE: 2000-10-30  
NUMBER OF SEQ ID NOS: 4385  
SOFTWARE: Friedmann macro Napro4  
SEQ ID NO 661  
LENGTH: 121  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-818-875-661

Query Match 2.2%; Score 121; DB 9; Length 121;  
Best Local Similarity 100.0%; Pred. No. 2.2e-20;  
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1. ATGATTATATGCTGCTCTTCCGCTTGAGAGTACAGTAAATGTCATTATGCTATGCGAAA 60  
|||||  
Db 1 ATGATTATATGCTGCTCTTCCGCTTGAGAGTACAGTAAATGTCATTATGCTATGCGAAA 60  
Qy 61 ATCTTAGAGTGTCCATGCTGCTGAGTGAATGATCAAGAACCTGCTCCACAAAGTGAC 120  
61 ATCTTAGAGTGTCCATGCTGCTGAGTGAATGATCAAGAACCTGCTCCACAAAGTGAC 120  
Db 121 C 121  
121 C 121

RESULT 12  
US-09-818-875-662/c  
Sequence 662, Application US/09818875  
Publication No. US20030051270A1  
GENERAL INFORMATION:  
APPLICANT: Kmiec, Eric B.  
APPLICANT: Gamper, Howard B.  
APPLICANT: Rice, Michael C.  
TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single  
TITLE OF INVENTION: Stranded Oligonucleotides  
FILE REFERENCE: Napro-4  
CURRENT APPLICATION NUMBER: US/09/818,875  
CURRENT FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: US 60/192,176  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/192,179  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: US 60/208,538  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: US 60/244,989  
PRIOR FILING DATE: 2000-10-30  
NUMBER OF SEQ ID NOS: 4385  
SOFTWARE: Friedmann macro Napro4  
SEQ ID NO 662  
LENGTH: 121  
TYPE: DNA

```

; ORGANISM: Homo sapiens
US-09-818-875-662

Query Match
Best Local Similarity 100.0%; Score 121; DB 9; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAA 60
    |||||||
DB 121 ATGGATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAA 62

QY 61 ATCTTAAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGAC 120
    |||||||
DB 61 ATCTTAAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGAC 2

QY 121 C 121
    |
DB 1 C 1

RESULT 13
US-09-818-875-665
; Sequence 665, Application US/09818875
; Publication No. US20030051270A1
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gamper, Howard B.
; APPLICANT: Rice, Michael C.
; TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single
; TITLE OF INVENTION: Stranded Oligonucleotides
; FILE REFERENCE: Napro-4
; CURRENT APPLICATION NUMBER: US/09/818,875
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/192,176
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/192,179
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/208,538
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: US 60/244,989
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 4385
; SOFTWARE: Friedman macro Napro4
; SEQ ID NO 665
; LENGTH: 121
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-818-875-665

Query Match
Best Local Similarity 100.0%; Score 121; DB 9; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAAATCT 64
    |||||||
DB 1 ATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAAATCT 60

QY 65 TAGAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGACACA 124
    |||||||
DB 61 TAGAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGACACA 120

QY 125 T 125
    |
DB 121 T 121

RESULT 14
US-09-818-875-666/C
; Sequence 666, Application US/09818875
; Publication No. US20030051270A1
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gamper, Howard B.
```

```

; APPLICANT: Rice, Michael C.
; TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single
; TITLE OF INVENTION: Stranded Oligonucleotides
; FILE REFERENCE: Napro-4
; CURRENT APPLICATION NUMBER: US/09/818,875
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/192,176
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/192,179
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/208,538
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: US 60/244,989
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 4385
; SOFTWARE: Friedman macro Napro4
; SEQ ID NO 666
; LENGTH: 121
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-818-875-666

Query Match
Best Local Similarity 100.0%; Score 121; DB 9; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAAATCT 64
    |||||||
DB 121 ATTATCTGCTCTTCGCGTTGAAGAAGTACAAATGTCATTATGCTATGCAGAAAATCT 62

QY 65 TAGAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGACACA 124
    |||||||
DB 61 TAGAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGTGTGACACA 2

QY 125 T 125
    |
DB 1 T 1

RESULT 15
US-09-818-875-669
; Sequence 669, Application US/09818875
; Publication No. US20030051270A1
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gamper, Howard B.
; APPLICANT: Rice, Michael C.
; TITLE OF INVENTION: Targeted Chromosomal Genomic Alterations with Modified Single
; TITLE OF INVENTION: Stranded Oligonucleotides
; FILE REFERENCE: Napro-4
; CURRENT APPLICATION NUMBER: US/09/818,875
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/192,176
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/192,179
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/208,538
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: US 60/244,989
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 4385
; SOFTWARE: Friedman macro Napro4
; SEQ ID NO 669
; LENGTH: 121
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-818-875-669

Query Match
Best Local Similarity 100.0%; Score 121; DB 9; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 AGAAATCTTAGAGTGTCCCATCTGCTGTGAGAGTTGATCAAGAACCTGTCTCCACAAAGT 115
```

```

Db      1  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OY      116 GTGACCATATTTTGCAGAAATTTGCGATGCTGAAGAACTCTCAACCCAGAGAAAGGGCCTT 175
Db      61  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
OY      176 C 176
Db      121 C 121

```

Search completed: June 27, 2003, 05:25:04  
 Job time : 736.151 secs